



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Reconhecimento de Cenários baseado nas
Localizações dos Fornecedores do Governo Federal**

Rodrigo Peres Ferreira

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. Marcelo Ladeira

Brasília
2018

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

PP437r Peres Ferreira, Rodrigo
Reconhecimento de Cenários baseado nas Localizações dos
Fornecedores do Governo Federal / Rodrigo Peres Ferreira;
orientador Marcelo Ladeira. -- Brasília, 2018.
72 p.

Tese (Doutorado - Mestrado Profissional em Computação
Aplicada) -- Universidade de Brasília, 2018.

1. Fornecedores de Fachada. 2. Reconhecimento de
Cenários. 3. Redes Neurais Convolucionais. 4. Deep
Learning. 5. Auditoria Governamental. I. Ladeira, Marcelo,
orient. II. Título.

Dedicatória

Dedico este trabalho a seus principais incentivadores e colaboradores: Prof Dr. Rommel Carvalho, que me introduziu ao mundo da Ciência de Dados e Aprendizagem de Máquina, ao meu orientador, Prof. Dr. Marcelo Ladeira, e a meu colega Jonas Wacker, por suas colaborações relevantes ao estudo.

Agradecimentos

Agradecimentos especiais à equipe do Observatório da Despesa Pública (ODP) do Ministério da Transparência e Controlaria-Geral da União (CGU) e ao Coordenador-Geral Leonardo Sales, pelo apoio institucional, pela divulgação do trabalho e pela paciência que tiveram em não poder contar comigo durante a reta final deste projeto, para que eu pudesse me concentrar na conclusão das análises.

Resumo

Uma das características mais comuns em qualquer esquema de corrupção ou desvio de recursos públicos é a presença de fornecedores de fachada, através dos quais se dá uma aparência legal a recursos com destinação ilícita. Este trabalho investiga a aplicação de um classificador de imagens para auxiliar na detecção de tais fornecedores, utilizando redes neurais convolucionais (CNN em inglês). Foram obtidas cerca de 120 mil imagens de 30 mil fornecedores com contratos ativos com o Executivo Federal entre 2016 e 2017, por meio de pesquisas de endereços pela API do Google Street View. Cada imagem baixada foi classificada entre cenários comerciais (lojas de rua) e não comerciais (diferentes de lojas de rua), através do ajuste fino de uma rede convolucional pré-treinada com milhões de imagens e centenas de categorias de cenários do banco de imagens Places365. Ao final, verificamos que é possível ajustar com sucesso modelos de redes convolucionais para diferenciar cenários entre comerciais e não comerciais, e o desempenho dos modelos ajustados não depende de orientações específicas das imagens. Além disso, os resultados obtidos com testes Qui-Quadrado e Regressão Logística permitiram verificar que existe uma relação estatisticamente significativa entre um cenário ser comercial e a diminuição de risco de punição de um fornecedor, especialmente em algumas classes de atividade econômica relacionadas à Administração Pública, Saúde e Outros Serviços. Entretanto, esta informação ainda possui baixo valor preditivo se comparada a algumas das tipologias já levantadas no âmbito do Ministério da Transparência e Controladoria-Geral da União (CGU).

Palavras-chave: Fornecedores de Fachada, Reconhecimento de Cenários, Redes Neurais Convolucionais

Abstract

In most corruption scandals, the use of front companies for money laundering is almost ubiquitous. This work investigates the application of an image classifier to detect such organizations, through the use of Convolutional Neural Networks (CNN). We collected about 120 thousand images from 30 thousand government suppliers contracted by the Executive Branch of the Brazilian Federal Government between 2016 and 2017, through address searching using the Google Street View API. Each image was classified between street market and non street market places, through fine-tuning a pretrained CNN with millions of images and hundreds of places categories from the Places365 database. Finally, we checked that it is possible to fine-tune convolutional neural networks to classify such scenes, and the performance of these networks is independent of image orientation. Our experiments results using Chi-Squared tests and Logistic Regression also showed us that exists a statistically significant relationship between places classified as street markets and lower risk for suppliers, especially for some economic sectors, such as those related to Public Administration, Health and Other Services. However, this information has low predictive value when compared to others already gathered by the Ministry of Transparency and Office of the Comptroller General (CGU, in Portuguese) auditors.

Keywords: Front companies, Scene Recognition, Convolutional Neural Networks

Sumário

1	Definição do Problema	1
1.1	A CGU e o Observatório da Despesa Pública	1
1.2	Justificativa do Tema	2
1.3	Hipóteses de Pesquisa	3
1.4	Objetivos	4
1.4.1	Objetivo Geral	4
1.4.2	Objetivos Específicos	4
1.5	Contribuição Esperada	4
1.6	Estrutura da Dissertação	5
2	Revisão Bibliográfica	6
2.1	Fundamentação Teórica	6
2.1.1	Fornecedores de Fachada	6
2.1.2	Imagens	7
2.1.3	Redes Neurais Artificiais	8
2.1.4	O algoritmo de <i>Backpropagation</i>	9
2.1.5	Redes Neurais Convolucionais (CNN)	10
2.1.6	<i>Transfer Learning</i>	12
2.1.7	Métricas de Classificação	13
2.2	Revisão do Estado da Arte	17
2.2.1	Detecção de Fraudes	17
2.2.2	Reconhecimento de Cenários	19
2.2.3	Arquiteturas ResNet	20
2.2.4	Bancos de Dados de Cenários	23
3	Solução Proposta	25
3.1	Metodologia	26
3.1.1	Obter as imagens dos fornecedores	27
3.1.2	Realizar anotações nas imagens obtidas	29

3.1.3	Ajustar modelos baseados em redes convolucionais	30
3.1.4	Construir modelos de risco gerais com diversos atributos	31
3.1.5	Realizar a validação dos modelos de risco gerais	33
4	Resultados	35
4.1	<i>Transfer Learning</i> da Rede Convolucional (CNN)	35
4.2	Comparando a saída da CNN com tipologias do ODP	37
4.3	Avaliando modelos gerais de risco de fornecedores	39
4.3.1	Modelo Base (Variáveis Binárias)	40
4.3.2	Modelo Quantitativo	43
4.3.3	Modelos de Regressão por CNAE	48
5	Conclusões e Trabalhos Futuros	51
	Referências	56

Lista de Figuras

2.1	Imagem com valores em escala de cinza (a) e a mesma imagem composta por valores nas escalas RGB em (b).	7
2.2	Exemplo de uma Rede Neural Artificial (ANN) simples, representada na figura por um grafo acíclico.	9
2.3	Exemplo de um fluxo <i>feedforward</i> de uma rede neural (a), em contraposição ao fluxo do algoritmo de aprendizado (<i>backpropagation</i>) em (b).	11
2.4	Arquitetura Genérica de uma rede neural convolucional.	11
2.5	Gráfico ROC, medindo o desempenho de cinco classificadores discretos (letras A a E).	15
2.6	Exemplo de Curva ROC gerada para diversos valores limite de um classificador probabilístico.	17
2.7	Exemplo de degradação no desempenho em redes com mais camadas (linha vermelha).	21
2.8	Exemplo de um bloco residual típico de uma arquitetura ResNet.	21
2.9	Exemplo de um bloco residual simples de uma ResNet e uma Wide ResNet.	22
2.10	Arquitetura da Wide Resnet 18.	23
3.1	Exemplo de requisições HTTP GET para a API do Street View para a obtenção de metadados e imagens.	28
3.2	Exemplo de imagens anotadas como áreas comerciais e não comerciais.	30
3.3	Exemplo de imagens classificadas pela rede convolucional AlexNet treinada no conjunto de dados Places205.	30
3.4	Exemplo de regiões que mais ativaram neurônios da penúltima camada da rede Wide ResNet18.	31
4.1	Abordagem utilizada para classificação dos cenários de fornecedores.	37
4.2	Curva ROC do modelo base.	43
4.3	Curva ROC do modelo quantitativo.	47
4.4	Curva ROC do modelo quantitativo sem a variável “prob_loja”.	47

5.1	Comparação de dois clusters de imagens de fornecedores baixadas do Google Street View utilizando as métricas de silhueta e distância máxima intra-cluster ($\text{Max}(a)$).	54
5.2	Exemplo de creche do programa ProInfância sendo construída ao longo dos anos.	55
5.3	Exemplo de rachadura visível em trecho de pista no centro de Brasília/DF.	55

Lista de Tabelas

2.1	Matriz de confusão para duas classes.	14
4.1	Classificação de imagens individuais: Média \pm Desvio Padrão AUC sobre 10 rodadas (com exceção do fine-tuning) para vários atributos.	36
4.2	Área sobre a curva ROC (AUC) da abordagem de Fine-Tuning da Wide Resnet18.	36
4.3	Campos derivados da saída da rede convolucional.	38
4.4	Relacionamentos entre Cenário Comercial e outras Tipologias.	39
4.5	Campos derivados das tipologias utilizadas para ajuste do modelo base.	40
4.6	Variáveis do modelo base.	42
4.7	Métricas de precisão, abrangência (<i>recall</i>) e F-Measure do modelo base.	42
4.8	Campos derivados das tipologias utilizadas para ajuste do modelo quantitativo.	44
4.9	Variáveis do modelo quantitativo.	45
4.10	Comparando o modelo com outras variáveis da rede convolucional.	46
4.11	Métricas de precisão, abrangência (<i>recall</i>) e F-Measure do modelo quantitativo.	46
4.12	Métricas de precisão, abrangência (<i>recall</i>) e F-Measure do modelo quantitativo sem a variável “prob_loja”.	47
4.13	Modelos de regressão com “prob_loja” por seção da CNAE.	49
4.14	Modelos de regressão com 12 variáveis do modelo quantitativo por seção da CNAE.	50

Lista de Abreviaturas e Siglas

AIC *Akaike Information Criteria.*

ANN Rede Neural Artificial.

AUC *Área sobre a Curva ROC.*

BN *Batch Normalization.*

BoVW *Bag-of-Visual-Words.*

CEIS *Cadastro de Entidades Inidôneas e Suspensas.*

CGU Ministério da Transparência e Controladoria-Geral da União.

CNN Redes Neurais Convolucionais.

Comprasnet SIASG Sistema de Administração de Serviços Gerais.

CRISP-DM *Cross Industry Standard Process for Data Mining.*

FNDE Fundo Nacional de Desenvolvimento da Educação.

ILSVRC *ImageNet Large Scale Visual Recognition Challenge.*

LDA Análise de Discriminantes Lineares.

MLP *Multilayer Perceptron.*

ODP Observatório da Despesa Pública.

ROC *Receiver operating characteristic.*

SPMK *Spatial Pyramid Matching Kernel.*

SVM Máquinas de Vetores de Suporte.

Capítulo 1

Definição do Problema

O estudo propõe-se a identificar fornecedores em locais comerciais e não comerciais e que possuam contratos com órgãos do Poder Executivo Federal, através da obtenção e classificação das imagens dos cenários onde estão localizados, e verificar se essa classificação pode melhorar ou complementar tipologias de risco de fornecedores já existentes no âmbito do Ministério da Transparência e Controladoria-Geral da União (CGU).

1.1 A CGU e o Observatório da Despesa Pública

O Ministério da Transparência e Controladoria-Geral da União (CGU) é o órgão central de controle interno do Poder Executivo Federal no Brasil. Conforme o Apêndice I, Art. 1º, inciso I do Decreto 8.910/2016 ¹, que define sua estrutura regimental, a CGU possui competência nos seguintes temas: Defesa do Patrimônio Público, Controle Interno, Auditoria Pública, Correição, Prevenção e Combate à Corrupção, Ouvidoria e Incremento da Transparência.

Para auxiliar a tomada de decisão dos auditores do órgão, bem como de gestores públicos de outros órgãos, a CGU dispõe de uma unidade especial denominada Observatório da Despesa Pública (ODP). O ODP é uma unidade de produção de informações estratégicas que realiza monitoramentos nas despesas do Governo Federal. Esse monitoramento é feito a partir de informações contidas em bases governamentais, utilizando técnicas de cruzamento e análise de dados [1].

O ODP produz dois tipos principais de informação: gerenciais e de auditoria. As informações gerenciais oferecem uma visão geral da distribuição de recursos, por área, ministério ou ano, por exemplo. Essas informações auxiliam gestores a identificar tendências em seus gastos, melhorando seus processos decisórios. As informações de auditoria

¹Decreto Nº 8.910 de 22 de novembro de 2016 (<http://www2.camara.leg.br/legin/fed/decret/2016/decreto-8910-22-novembro-2016-783927-publicacaooriginal-151417-pe.html>).

apresentam indícios de irregularidades ou mau uso de recursos públicos. Essas informações auxiliam os auditores do órgão a priorizar suas ações de fiscalização. Um exemplo deste tipo de informação são os alertas gerados a partir de cruzamentos de dados, como: servidores que possuem vínculo com licitantes ou fornecedores do governo, gastos de cartões corporativos em estabelecimentos atípicos ou licitantes com sócios em comum que participam de um mesmo processo licitatório.

Dentre os temas monitorados pelo ODP, as licitações públicas ocupam um papel relevante. Licitações totalizando cerca de R\$ 170 bilhões foram registradas nos últimos dois anos no Sistema de Administração de Serviços Gerais (Comprasnet SIASG), o principal sistema de licitações públicas do país². Durante estudos de monitoramento neste tema, o ODP costuma coletar imagens de sedes de companhias sob suspeita decorrentes dos alertas gerados. Uma forma de coletar as imagens é através da pesquisa de endereço pelas ferramentas Google Maps ou Google Street View. Neste procedimento, já foram encontrados cenários suspeitos, como áreas rurais ou lugares isolados.

1.2 Justificativa do Tema

Atualmente, sabe-se que a corrupção e a utilização de corporações de fachada são tópicos intimamente relacionados. De acordo com o Relatório da Iniciativa de Recuperação de Ativos Roubados do Banco Mundial e das Nações Unidas [2], a utilização deste subterfúgio é o método mais comum utilizado para esconder recebimento de fundos ou ativos recebidos ilegalmente.

Fato este reforçado pela Operação Lava Jato da Polícia Federal, que, desde março de 2014, ao investigar a relação entre doleiros e diretores da Petrobras, vem desvendando um mega-esquema de corrupção envolvendo autoridades ocupantes dos mais altos cargos públicos do país. Ainda em 2016, com base em dados do Ministério Público, Polícia Federal e Receita Federal, o jornal Estado de São Paulo³ identificou 34 empresas investigadas pela operação consideradas de fachada ou “noteiras”⁴, que receberam movimentação aproximada de R\$ 2,5 bilhões, provenientes de outras empresas com contratos com o governo. No âmbito da CGU, algumas informações úteis para a identificação destes fornecedores são levantados periodicamente, tais como: baixa quantidade de funcionários, baixa qualificação

²De acordo com as informações disponíveis no Comprasnet SIASG, em junho de 2018. Itens com valores maiores ou iguais a R\$ 1 bilhão (possíveis erros de preenchimento) foram excluídos.

³‘Estado’ mapeia 34 firmas suspeitas de atuação fictícia e identificadas em 2 anos e 4 meses de operação; em sua maioria, recursos foram repassados por empreiteiras (<http://politica.estadao.com.br/noticias/geral,empresas-de-fachada-na-lava-jato-giram-r-2-5-bi,10000067543>).

⁴Empresa “noteira”, em nosso contexto, é sinônimo de empresa de fachada, ou seja, criada com a intenção de fornecer aparência de legalidade na prestação de um bem ou serviço: legal apenas na “nota fiscal” apresentada.

e remuneração dos sócios, microempresas ou empresas de pequeno porte com licitações de valores incompatíveis, mais de um fornecedor no mesmo endereço, grande quantidade de classes de atividade econômica apresentadas pelo fornecedor, diferença entre data de abertura do fornecedor e da compra pela Administração inferior a 30 dias, fornecedores com penalidades suspensivas no cadastro, fornecedores no Cadastro de Entidades Inidôneas e Suspensas (CEIS), fornecedores com sócios que possuem empresas no CEIS e produto ou serviço fornecido incompatível com a atividade econômica do fornecedor.

Algumas destas informações também servem de insumo para um modelo interno à CGU para ranqueamento de contratos governamentais com base em critérios de risco [3] (mais detalhes no Capítulo 2), que utiliza dados da capacidade operacional e do histórico de contratos dos fornecedores como um de seus componentes.

Entendemos que, embora úteis, as informações levantadas por sistemas e modelos existentes no órgão podem ser complementadas. Dados importantes, muitas vezes, encontram-se em fontes não estruturadas, tais como textos ou imagens. Estas fontes de dados não são utilizadas de forma automatizada pelos modelos do órgão. Acreditamos que a classificação automática de imagens voltadas à identificação de cenários atípicos para os fornecedores é uma fonte valiosa de informação que pode complementar as informações já levantadas por sistemas e modelos da CGU, contribuindo para um aperfeiçoamento na detecção de fornecedores de fachada.

1.3 Hipóteses de Pesquisa

As principais hipóteses a serem verificadas pelo trabalho são:

- **com base na análise automatizada da imagem da sede de uma empresa é possível classificá-la entre cenários comerciais e não comerciais.** Pretende-se verificar que modelos de redes neurais convolucionais podem ser úteis para identificar padrões nas imagens de fornecedores que indiquem características dos locais onde estão localizados. Neste trabalho, estas características permitirão diferenciar os locais identificados entre cenários comerciais e não comerciais.
- **a identificação de empresas em cenários comerciais e não comerciais é um indicador relevante para determinar se ela é uma empresa de fachada ou um fornecedor de risco.** Pretende-se verificar se a classificação de cenários pode gerar um atributo relevante para que, juntamente com outros atributos já obtidos pela CGU (como, por exemplo, baixa quantidade de empregados e baixo nível de escolaridade dos sócios), seja possível melhorar o trabalho dos auditores na identificação dessas empresas.

1.4 Objetivos

Nesta seção, levantamos os objetivos geral e específicos a serem atingidos de forma a contribuir nos trabalhos realizados no âmbito da CGU.

1.4.1 Objetivo Geral

Avaliar, no âmbito da CGU, os métodos para identificação de fornecedores de fachada e fornecedores de risco através da classificação e análise de imagens de sedes de fornecedores contratados pelo Governo Federal para o fornecimento de determinados produtos ou serviços.

1.4.2 Objetivos Específicos

- ajustar modelos de Redes Neurais Convolucionais (CNN) existentes para classificação de cenários de forma a torná-los capazes de identificar de forma automatizada padrões em cenários de sedes de fornecedores do Governo Federal.
- verificar se houve melhora na identificação de fornecedores de fachada ou fornecedores de risco através da inclusão da classificação automática de imagens quando comparada a abordagens anteriores que não utilizavam esta informação.

1.5 Contribuição Esperada

Como contribuição esperada após o fim do projeto, espera-se introduzir o uso de imagens como fonte de dados em outros projetos no âmbito da CGU, como por exemplo no acompanhamento de obras públicas e na detecção de pessoas utilizadas como sócios simbólicos de empresas de fachada (possíveis “laranjas”) através de seus respectivos endereços.

Cabe ressaltar que a CGU atualmente não utiliza imagens de forma automatizada como insumo para suas ações de auditoria. Os dados utilizados são normalmente provenientes de bancos de dados relacionais. Portanto, a criação de um classificador de imagens representa, no contexto do órgão, uma inovação nas tecnologias utilizadas para apoio ao trabalho dos auditores.

1.6 Estrutura da Dissertação

Após este capítulo introdutório, no Capítulo 2 (Revisão Bibliográfica) revisamos os conceitos mais importantes para a compreensão deste trabalho e a evolução do estado da arte em Detecção de Fraudes utilizando Mineração de Dados, Reconhecimento de Cenários e Redes Neurais Convolucionais.

No Capítulo 3 (Solução Proposta), abordamos as principais etapas a serem executadas ao longo deste trabalho, a fim de atingirmos os objetivos geral e específicos mostrados neste capítulo.

No Capítulo 4 (Resultados), apresentamos resultados comparativos de técnicas de ajuste de uma rede convolucional pré-treinada para tornar o classificador capaz de distinguir entre cenários comerciais e não comerciais. Além disso, utilizamos testes Qui-Quadrado e Regressões Logísticas a fim de verificar se a classificação dada pela rede convolucional é uma informação relevante junto a outros atributos comumente levantados pelos auditores da CGU.

No Capítulo 5 (Conclusões e Trabalhos Futuros), discutimos de forma resumida os resultados encontrados, com proposta de melhoria para trabalhos futuros, bem como fazemos referência a projetos relacionados que terão início no âmbito da CGU.

Todas as imagens e tabelas que referenciamos nas próximas páginas possuem indicação de fonte, com exceção das de autoria própria, criadas em consequência das pesquisas realizadas neste projeto.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, apresentamos uma revisão bibliográfica levando em conta conceitos importantes para a compreensão deste trabalho e a evolução do estado da arte nas áreas de Detecção de Fraudes, Reconhecimento de Cenários, Arquiteturas de Redes Convolucionais e Bancos de Dados de Cenários.

2.1 Fundamentação Teórica

Nesta seção, discutimos conceitos importantes para o entendimento deste trabalho, tais como: Fornecedores de Fachada, Imagens, Redes Neurais Artificiais, *Backpropagation*, Redes Neurais Convolucionais, *Transfer Learning* e Métricas de Classificação.

2.1.1 Fornecedores de Fachada

De acordo com Willebois et al. [2], as organizações ou fornecedores de fachada são tipos de corporações criadas e utilizadas para esconder identidades de pessoas envolvidas em casos de corrupção e lavagem de dinheiro.

As organizações de fachada podem ser operacionais ou não operacionais. As organizações não operacionais não possuem operações próprias, ativos, negócios significativos ou empregados ativos. Podem ser criadas com propósitos lícitos, como para facilitar a fusão de duas empresas. Entretanto, são comumente utilizadas por bancos privados para disfarçar a real propriedade dos ativos de seus clientes, dificultando o controle das autoridades. As organizações operacionais possuem um fluxo de operações ativo, permitindo que ativos ilícitos sejam misturados a fundos lícitos, facilitando operações de lavagem de dinheiro. Normalmente é o tipo mais envolvido em operações de entrega e recebimento de propina.

Dos tipos descritos, o escopo deste trabalho encontra-se melhor relacionado às organizações operacionais, que possuem operações ativas e portanto podem ser encontradas a céu aberto, permitindo a utilização de informações dos cenários das sedes dessas entidades. As organizações não operacionais comumente existem sem a necessidade de criação de locais físicos.

2.1.2 Imagens

Conforme Qidwai e Chen [4], uma imagem (bidimensional, ver Figura 2.1 (a)) é uma função de duas variáveis $a(x, y)$ onde a é a amplitude da imagem nas suas coordenadas (x, y) . As coordenadas e a amplitude podem ser números reais ou inteiros. No primeiro caso, a imagem é considerada analógica, enquanto no segundo temos uma imagem digital. A imagem digital é geralmente resultante de um processo de quantização da imagem analógica, que converte os valores contínuos reais em números inteiros discretos. No contexto de processamento de imagens, chamamos as coordenadas (x, y) de **pixels** e o valor da amplitude $a(x, y)$ de **brilho**.

Esta definição de imagens bidimensionais (normalmente compostas de valores de brilho em escala de cinza) pode ser facilmente estendida para imagens tridimensionais (Figura 2.1 (b)), compostas por valores em três escalas de cores (vermelho, verde e azul, ou “RGB” em inglês), se para cada pixel (x, y) adicionarmos como entrada a variável z correspondente a uma das escalas (ou canais) de cor utilizadas. Desta forma, a imagem tridimensional colorida seria a função $a(x, y, z)$ onde a é o brilho da imagem no pixel (x, y) do canal z .

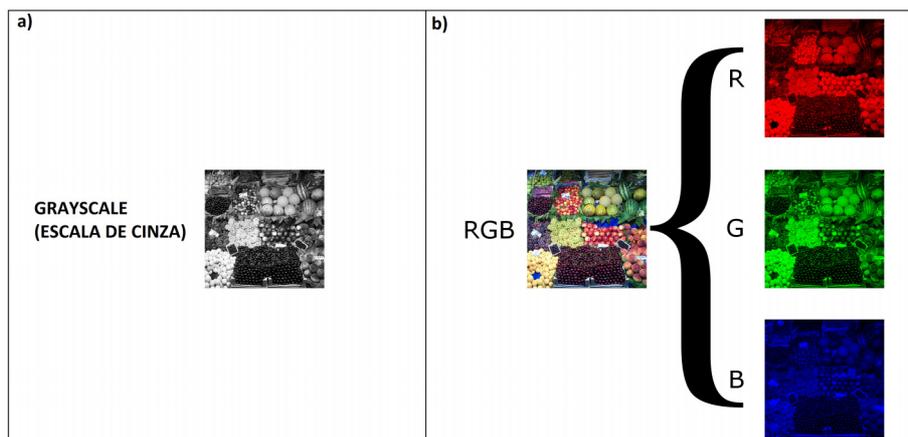


Figura 2.1: Imagem com valores em escala de cinza (a) e a mesma imagem composta por valores nas escalas RGB em (b).

Fonte: <https://commons.wikimedia.org/w/index.php?curid=18673639>

As imagens de fornecedores que obtivemos do Google Street View são imagens tridimensionais coloridas nas escalas RGB.

2.1.3 Redes Neurais Artificiais

Conforme LeCun, Bengio e Hinton [5], muitas aplicações de aprendizagem de máquina utilizam classificadores lineares. Normalmente, estes classificadores calculam somas ponderadas de atributos: se o valor desta soma for superior a determinado limite, a entrada é classificada como sendo de uma determinada categoria.

Entretanto, nem todos os tipos de dados possuem padrões simples. Dados provenientes de imagens exigem funções de decisão mais complexas, capazes de lidar com o dilema invariância versus seletividade, que consiste em ignorar padrões irrelevantes na imagem (diferenças de posição e iluminação, por exemplo), porém mantendo a sensibilidade para considerar padrões pequenos mas relevantes para a classificação (diferença entre um cão e um lobo, por exemplo). Um classificador linear simples treinado a partir de pixels brutos pode colocar dois cães de mesma raça em categorias diferentes com base nas diferentes posições e iluminações dessas imagens, ao mesmo tempo em que poderia colocar um cão e um lobo dentro de uma mesma categoria caso estejam em posições e condições de iluminação semelhantes.

Uma das soluções para esse dilema é fazer uma seleção cuidadosa de atributos que permitam distinguir as categorias em questão. Uma das principais vantagens do uso de Redes Neurais Artificiais profundas é que elas permitem a extração de atributos úteis de forma automática a partir do próprio processo de treinamento do modelo.

Vamos definir uma **Rede Neural Artificial** (em inglês, *Artificial Neural Network* - ANN) pelo seu exemplo mais comum, as *Feedforward Neural Networks*, também chamadas de *Multilayer Perceptrons (MLP)*. De acordo com Goodfellow, Bengio e Courville [6], uma Rede Neural Artificial do tipo *feedforward* define um mapeamento $y = f(x; \theta)$ e aprende os parâmetros θ a fim de aproximar y ao valor da categoria especificada para uma dada entrada x . São chamadas de redes *feedforward* já que as computações fluem de forma unidirecional da entrada x até a saída y . Não existem portanto conexões de saída retroalimentando as entradas do modelo, de modo que podemos representar esses modelos através de um grafo acíclico como o representado na Figura 2.2.

Conforme Goodfellow, Bengio e Courville [6], estes modelos são chamados de **redes** justamente por serem representados por uma composição de funções, como, por exemplo, $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$, onde $f^{(1)}$ representa a primeira camada, $f^{(2)}$ a segunda camada e assim por diante. As camadas intermediárias da rede são chamadas de camadas **escondidas**, enquanto a última camada é chamada de camada de **saída**.

Cada uma das camadas é representada por um vetor, e cada elemento desse vetor é chamado de “neurônio”. De acordo com Goodfellow, Bengio e Courville [6], o termo **neural** desses modelos vem do fato de que foram levemente inspirados pela neurociência. O valor de cada neurônio do vetor é resultado da aplicação de uma função $f^{(i)}(x)$ aos

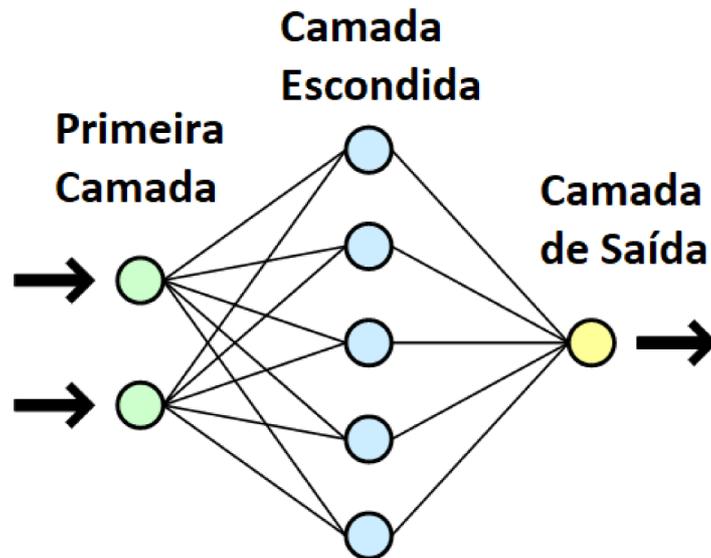


Figura 2.2: Exemplo de uma Rede Neural Artificial (ANN) simples, representada na figura por um grafo acíclico.

Fonte: <https://commons.wikimedia.org/w/index.php?curid=1412126>

valores da camada anterior. Esta função, chamada função de ativação, também possui inspiração na neurociência, e pode ser implementada de diversas formas: através das funções logística ($f(x) = \frac{1}{1+e^{-x}}$) ou tangente hiperbólica ($\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$) ou ainda, em redes mais modernas, da função ReLU (*Rectified Linear Unit*), que possui a forma $f(x) = \max(0, x)$.

Conforme Goodfellow, Bengio e Courville [6], embora parte das características de uma rede neural seja inspirada biologicamente pelo ramo da neurociência, devemos levar em consideração que muitas das características das redes mais modernas são guiadas por questões matemáticas e de engenharia, de forma que seria melhor entendermos esses modelos como aproximadores de funções, e não propriamente reproduções fiéis de um cérebro biológico.

2.1.4 O algoritmo de *Backpropagation*

O algoritmo de *Backpropagation* é o algoritmo utilizado para ajustar os parâmetros da rede (θ) de forma a torná-la uma boa função de aproximação das categorias desejadas. É através dele que a rede aprende a diferenciar quais dos atributos de entrada são relevantes para a classificação.

Vamos considerar um conjunto de treinamento que possui determinados valores de

entrada e um vetor de categorias de saída. Conforme Grus [7], o algoritmo de *Backpropagation* pode ser descrito em cinco etapas:

1. Executar uma passagem pela rede da entrada para a saída (*feedforward*) com os parâmetros iniciais;
2. Obter o erro de saída da rede (isto é, a diferença entre o vetor de saída da última camada da rede e o vetor de categorias do conjunto de treinamento);
3. Computar o gradiente (ou derivada) do erro de saída em função dos parâmetros dos neurônios da última camada, e ajustar estes parâmetros na direção que mais diminui esta função de erro;
4. Computar o gradiente (ou derivada) do erro de saída em função dos parâmetros dos neurônios das camadas intermediárias e iniciais a partir dos gradientes das últimas camadas através da regra da cadeia¹;
5. Ajustar os parâmetros das camadas intermediárias e iniciais na direção que mais diminui o erro a partir dos gradientes calculados.

Tipicamente, estas cinco etapas são executadas diversas vezes até a convergência da rede, ou seja, até que o valor do erro da rede não diminua de forma significativa.

O algoritmo é chamado de *Backpropagation* devido ao fato de que, no passo 4, os gradientes para as camadas intermediárias e iniciais da rede são calculados a partir dos gradientes da última camada, de modo que o processamento para computação desses gradientes é feito de forma inversa (de trás para frente) como ilustra a Figura 2.3. Isso ocorre como consequência da aplicação da regra da cadeia, onde primeiro são calculadas as derivadas dos parâmetros de saída do modelo (função mais externa) para depois serem calculadas as derivadas dos parâmetros das camadas intermediárias (funções internas).

2.1.5 Redes Neurais Convolucionais (CNN)

Uma rede neural convolucional (ver Figura 2.4) se diferencia de outras arquiteturas principalmente pela adição de camadas iniciais que realizam algumas operações específicas. De acordo com LeCun, Bengio e Hinton [5], essas operações visam tirar proveito das propriedades de sinais naturais (tais como as imagens). As ideias principais por trás dessas operações são: conexões locais, parâmetros compartilhados, subamostragens e o uso de múltiplas camadas.

¹A regra da cadeia do cálculo de derivadas nos diz que a derivada de uma função composta (como uma rede neural) $F(x) = f(g(x))$ é calculada multiplicando a derivada da função mais externa (de saída) pela derivada da função mais interna: $F'(x) = f'(g(x))g'(x)$.

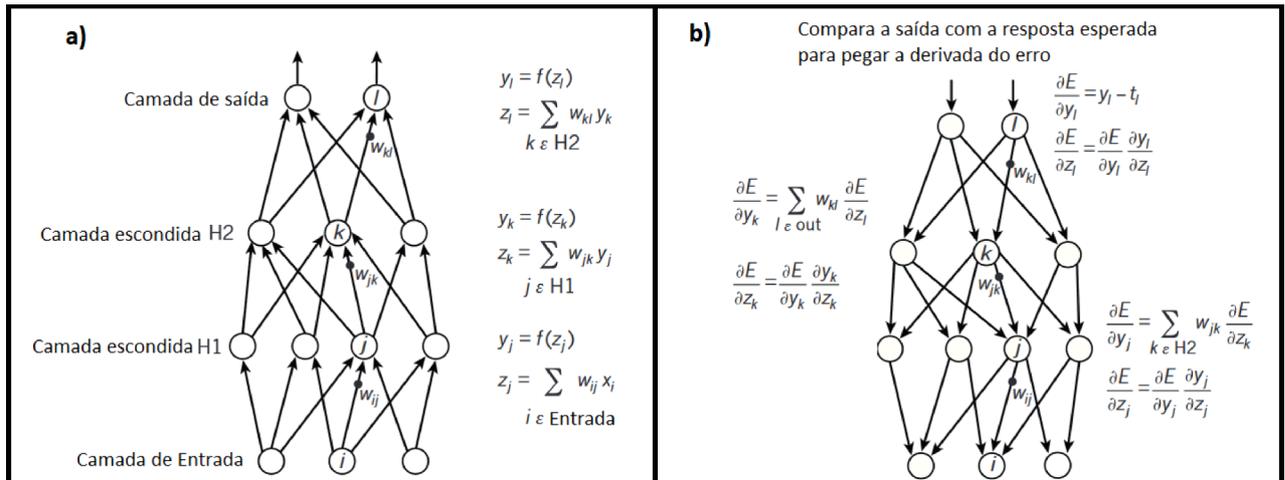


Figura 2.3: Exemplo de um fluxo *feedforward* de uma rede neural (a), em contraposição ao fluxo do algoritmo de aprendizado (*backpropagation*) em (b).

Fonte: Lecun, Bengio e Hinton (2015) [5].

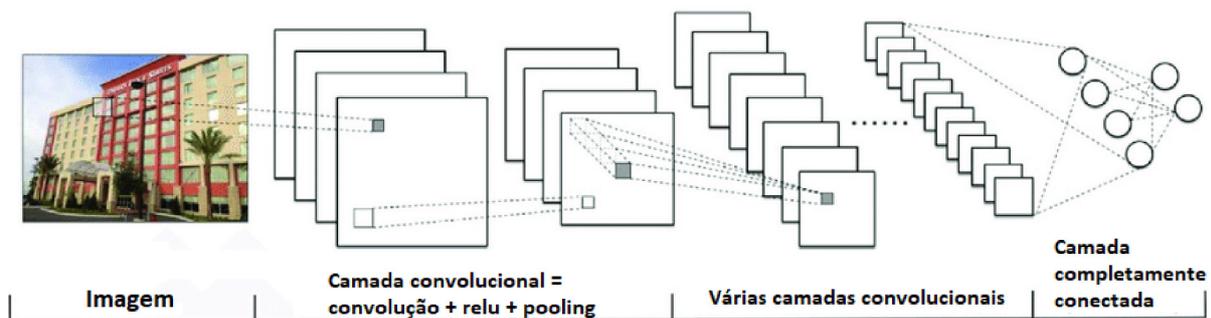


Figura 2.4: Arquitetura Genérica de uma rede neural convolucional.

Fonte: Ma et al. (2018) [8].

Uma dessas operações é chamada de **convolução**, que divide a imagem em janelas menores e as processa em conjuntos de unidades locais da rede chamados de *feature maps*, com o objetivo de recuperar elementos similares que apareçam em diferentes lugares de uma mesma imagem sem aumentar excessivamente a quantidade de parâmetros a serem treinados. Para isso, a operação de convolução deve compartilhar os mesmos parâmetros (ou pesos) multiplicativos, que em conjunto formam um **filtro** para cada *feature map* em todas as partes da imagem. Na Figura 2.4 podemos ver como as operações de convolução são aplicadas em regiões específicas, alterando também as dimensões da imagem de entrada ao longo dessas camadas.

Entre operações de convolução e funções de ativação ReLU, é aplicada outra operação importante na maioria das redes convolucionais: o *Pooling*. O *Pooling* é uma operação de subamostragem que visa reduzir as dimensões da imagem, facilitando a junção de padrões

comuns e tornando o classificador resistente a pequenas translações ou ruídos locais. Esta operação também diminui a quantidade de parâmetros a serem treinados.

Arquiteturas de redes convolucionais mais modernas também aplicam técnicas de normalização de dados ao longo da rede, chamadas de *Batch Normalization* (BN). De acordo com Ioffe e Szegedy [9], a utilização de *Batch Normalization* acelera o ajuste dos parâmetros da rede e ajuda a prevenir seu *overfitting*.

As operações de convolução e *pooling* funcionam de forma análoga à seleção manual de atributos dos algoritmos de aprendizagem de máquina tradicionais. Após a execução destas camadas, a rede deve ter sido capaz de identificar os padrões relevantes para classificação dos cenários, descartando os menos relevantes. A partir deste ponto, são aplicadas camadas de rede completamente conectadas às camadas anteriores, similares às redes *feedforward* mais comuns, que se encarregarão do trabalho de classificação em si.

Embora a seleção de atributos possa ser facilitada com o uso dessas redes, há um trabalho adicional na seleção dos chamados hiperparâmetros. Bengio [10] define hiperparâmetro da seguinte forma: “Um hiperparâmetro de um algoritmo de aprendizagem A é uma variável configurada antes da aplicação de A nos dados que não seja selecionada diretamente pelo próprio algoritmo”. Para o caso de modelos de redes neurais profundas, alguns exemplos de hiperparâmetros são: tipo de função de custo a ser otimizada pelo algoritmo de aprendizagem (*Loss Function*), taxa de aprendizagem (*Learning Rate*), que é a taxa com que os parâmetros do modelo são ajustados para minimizar a função de custo a cada iteração do algoritmo de *Backpropagation*, quantidade de camadas intermediárias (*Number of Hidden Layers*) e quantidade de neurônios nas camadas intermediárias da rede (*Number of Hidden Units*).

2.1.6 *Transfer Learning*

Um dos objetivos deste trabalho é ajustar modelos de Redes Neurais Convolucionais (CNN) para torná-los capazes de reconhecer cenários comerciais e não comerciais, e a forma mais comum de se ajustar um modelo de classificação baseado em redes convolucionais é através de técnicas de Transferência de Aprendizado (*Transfer Learning*).

Conforme Yosinski et al. [11], *Transfer Learning* é um conjunto de técnicas que se baseiam no treinamento de uma rede base que terá seus atributos reutilizados (ou transferidos) para o treinamento de uma rede alvo. O processo tende a funcionar melhor quanto mais gerais forem os atributos aprendidos pela rede base. Existem duas formas básicas de realizar *Transfer Learning*:

1. utilizando a rede convolucional como um extrator de atributos: consiste em extrair a saída de uma das camadas intermediárias da rede e utilizar esta saída para trei-

nar um modelo externo em um contexto diferente, como uma Máquina de Vetores Suporte (SVM) ou Regressão Logística.

2. ajuste fino (*fine tuning*): consiste em extrair a camada final da rede e substituí-la por outra de contexto diferente, continuando o treinamento da rede a partir dessa substituição. Normalmente, os pesos das camadas iniciais da rede são fixados para preservar as representações mais gerais criadas pelo modelo anterior.

As duas abordagens são muito populares e possuem a vantagem de permitir que sejam aproveitados os atributos já aprendidos pelo modelo, o que facilita o processo de treino e permite que sejam obtidos bons resultados mesmo com uma quantidade pequena de imagens de treino (contanto que sejam semelhantes o bastante com o conjunto original utilizado para treinar o modelo do início).

Neste trabalho, ajustamos uma rede convolucional pré-treinada em um conjunto de imagens com milhões de cenários para torná-lo capaz de diferenciar cenários comerciais de não comerciais. Para ajustá-lo, utilizamos as duas abordagens descritas nesta seção: extração de atributos para treino de um modelo externo (SVM) e também ajuste fino da rede substituindo a última camada por uma nova camada contendo apenas duas categorias (“loja” e “não loja”).

Os modelos serão avaliados utilizando métricas de classificação comuns em aprendizagem de máquina, explicadas na próxima subseção.

2.1.7 Métricas de Classificação

Para avaliarmos os modelos ajustados por *Transfer Learning* neste trabalho, bem como outros modelos de risco de fornecedores, utilizamos algumas das métricas mais utilizadas em Aprendizagem de Máquina. Estas métricas são baseadas na Matriz de Confusão, que consiste em uma comparação de quantidades entre valores reais e previstos pelo modelo de classificação.

Para modelos de classificação binária (apenas duas classes, positiva e negativa), esta matriz assume a forma descrita na Tabela 2.1. As linhas da matriz correspondem às classes reais e as colunas, às classes previstas pelo modelo. Desta forma, a célula correspondente à primeira linha e à primeira coluna computam os valores de **Verdadeiros Positivos**, ou seja, quantos dados marcados como positivos de fato são previstos como positivos pelo modelo. Já célula da primeira linha e segunda coluna correspondem a valores de **Falsos Negativos**, indicando quantos dados marcados como positivos foram erroneamente marcados como negativos pelo modelo. A segunda linha e primeira coluna correspondem a valores de **Falsos Positivos**, indicando dados marcados como negativos mas que foram classificados erroneamente como positivos. Por fim, a segunda linha e

segunda coluna correspondem a valores de **Verdadeiros Negativos**, indicando quantos dados marcados como negativos são corretamente classificados como negativos.

Tabela 2.1: Matriz de confusão para duas classes.

		Classe prevista	
		Positivo	Negativo
Classe real	Positivo	Verdadeiros Positivos	Falsos Negativos
	Negativo	Falsos Positivos	Verdadeiros Negativos

A partir da Matriz de Confusão, são definidas as métricas de Acurácia, Precisão e Abrangência (ou *Recall*).

A Acurácia é a porcentagem total de acertos de um modelo de classificação, ou seja, a taxa de Verdadeiros Positivos e Verdadeiros Negativos sobre os valores totais da matriz de confusão, conforme a equação a seguir:

$$Acurácia = \frac{VerdadeirosPositivos(VP) + VerdadeirosNegativos(VN)}{Total}$$

A Precisão é a porcentagem de acertos que um modelo classificador possui ao prever a classe positiva, ou seja, a taxa de Verdadeiros Positivos sobre a soma de Verdadeiros Positivos e Falsos Positivos, conforme a equação:

$$Precisão = \frac{VerdadeirosPositivos(VP)}{VerdadeirosPositivos(VP) + FalsosPositivos(FP)}$$

Já a Abrangência (ou *Recall*) é a porcentagem de acertos que um classificador possui na classe positiva, ou seja, quantos elementos da classe positiva o classificador é capaz de detectar (equivalente à taxa de verdadeiros positivos). É dados pela equação:

$$Recall = \frac{VerdadeirosPositivos(VP)}{VerdadeirosPositivos(VP) + FalsosNegativos(FN)}$$

Uma métrica derivada de Precisão e *Recall* é a Medida F1, ou *F-Measure*, dada pela média harmônica entre as duas medidas anteriores:

$$F_1 = 2 \cdot \frac{Precisão \cdot Recall}{Precisão + Recall}$$

Desta forma, a *F-Measure* procura ser uma medida de equilíbrio entre Precisão e *Recall*.

Também derivada da matriz de confusão, temos a medida de Área sobre a Curva ROC (AUC). Mas para definirmos a AUC, temos primeiro que definir o que é um gráfico ROC (*Receiving Operating Characteristic*). De acordo com Fawcett [12], o gráfico ROC é um gráfico bidimensional onde a taxa de verdadeiros positivos (tp = Verdadeiros Positivos/Total de Positivos, o mesmo que *Recall*) é representada no eixo Y e a taxa de falsos positivos (fp = Falsos Positivos/Total de Negativos) é representada no eixo X. Intuitivamente, podemos compreender o gráfico ROC como um gráfico de compromisso entre os custos (Falsos Positivos) e benefícios (Verdadeiros Positivos) de um modelo classificador.

Um classificador discreto é aquele que possui apenas um rótulo como resposta (verdadeiro ou falso, positivo ou negativo). Ele é representado no gráfico ROC como apenas um ponto no plano, no par de coordenadas (tp,fp). A Figura 2.5 representa o plano do gráfico ROC com resultados de cinco classificadores discretos (de A a E).

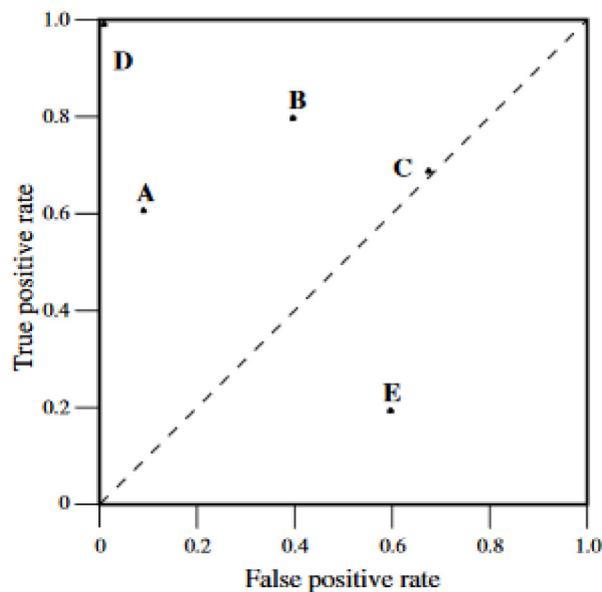


Figura 2.5: Gráfico ROC, medindo o desempenho de cinco classificadores discretos (letras A a E).

Fonte: Fawcett (2006) [12].

Desta forma, o que queremos é um classificador com maior custo-benefício possível, próximo do ponto (1,0) do gráfico. Ou seja, um classificador com taxa de verdadeiros positivos próxima de 100% e mantendo uma taxa de falsos positivos próxima de zero.

Observe que a linha pontilhada na diagonal do gráfico da Figura 2.5 representa a região dos classificadores aleatórios. Se tivéssemos um classificador discreto na coordenada (0.5,0.5), isso significaria um classificador que acerta a classe positiva metade das vezes, o que seria o esperado de um classificador perfeitamente randômico. O mesmo ocorre com

qualquer outro ponto da linha pontilhada. Se um classificador aleatório acertar a classe positiva 70% das vezes, é esperado que sua taxa de falsos positivos também seja de 70% (como é o caso do classificador C).

Pontos na parte inferior direita da linha pontilhada representam em teoria classificadores piores que classificadores aleatórios. Entretanto, basta negarmos os resultados destes classificadores para obtermos classificadores equivalentes na parte superior esquerda da linha (veja o caso do classificador B na Figura 2.5, que é apenas a negação do classificador E). Desta forma, de acordo com Fawcett [12], classificadores nesta região não são de fato piores que os aleatórios, mas sua informação está sendo aplicada de forma incorreta.

É importante observar que redes neurais não são classificadores discretos. De fato, redes neurais retornam valores de probabilidades (ex. probabilidade de um elemento pertencer à determinada categoria) em vez da categoria em si. Para analisarmos este tipo de classificador utilizando um gráfico ROC é necessário utilizar um valor limite para produzir pontos no gráfico. Por exemplo, se a probabilidade da categoria for superior a um valor limite estabelecido, então o classificador produz como resultado o valor correspondente a aquela categoria. Desta forma, é possível discretizar o classificador probabilístico e produzir um ponto de taxa de verdadeiros positivos versus taxa de falsos positivos no gráfico.

Normalmente, ao avaliarmos um classificador probabilístico, testamos diversos valores limite, e portanto teremos diversos pontos com diferentes taxas de verdadeiros positivos e falsos positivos para cada valor limite estabelecido, perfazendo uma função em forma de escada, como a mostrada pela Figura 2.6. A medida em que aumentamos a quantidade de limites testados, a função em forma de escada se aproximará do formato de uma curva contínua.

Conforme Fawcett [12], a Curva ROC é uma representação bidimensional do desempenho de um modelo classificador. Se quisermos um valor escalar para avaliarmos o modelo, podemos calcular a **Área sobre a Curva ROC (AUC)**. Como a AUC está contida em uma porção de área de gráfico com dimensões 1x1, temos que ela sempre possuirá um valor entre 0 e 1. Como nenhum classificador pode ter realisticamente uma área inferior à diagonal do gráfico ROC, normalmente encontraremos a AUC entre valores de 0.5 (equivalente a um classificador randômico) a 1 (classificador com 100% de verdadeiros positivos e 0% de falsos positivos).

De acordo com Fawcett [12], a AUC tem uma propriedade estatística interessante: ela equivale à probabilidade de um classificador gerar um *score* de classificação maior para um elemento da classe positiva do que o *score* gerado para a classe negativa.

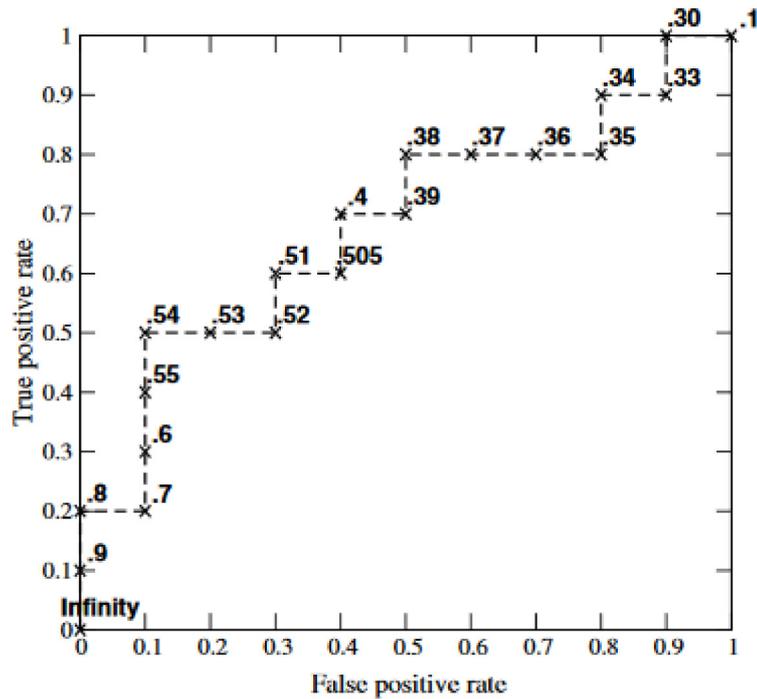


Figura 2.6: Exemplo de Curva ROC gerada para diversos valores limite de um classificador probabilístico.

Fonte: Fawcett (2006) [12].

2.2 Revisão do Estado da Arte

Nesta seção, discutiremos sobre a evolução do estado da arte nas áreas de pesquisa ligadas a este projeto, como Detecção de Fraudes, Reconhecimento de Cenários, Arquiteturas de Redes Convolucionais modernas (Resnet e Wide Resnet) e Bancos de Dados de Cenários.

2.2.1 Detecção de Fraudes

A detecção de fraudes através de técnicas de mineração de dados, assim como técnicas de detecção de fornecedores de fachada em específico, foi largamente influenciada pelos trabalhos de Fawcett e Provost em [13]. Os autores propuseram a utilização de aprendizagem de máquina na aplicação de um programa que descobre regras a partir de comportamentos fraudulentos de uma grande base de dados de transações telefônicas. A metodologia foi aplicada para detecção de clonagem de celulares, e os resultados indicam que esta abordagem é superior a métodos manuais de detecção de fraudes, além de prover adaptação a mudanças, tão comuns em ambientes onde ocorrem fraudes.

Neville e Jensen [14] propuseram a utilização incremental de classificadores bayesianos a partir de dados com alto grau de relacionamento, utilizando como exemplo a identifi-

cação de operações de lavagem de dinheiro e empresas de fachada através de dados de depósitos bancários e redes de relacionamentos.

Outras estratégias de detecção de fraudes através de classificação supervisionada incluem, de acordo com Bolton e Hand [15]: Análise de Discriminantes Lineares (LDA), Regressão Logística, *Random Forest* e Redes Neurais, além de combinações desses modelos.

Há ainda a aplicação de técnicas não supervisionadas, quando não há registros anteriores de comportamentos fraudulentos. De acordo com Bolton e Hand [15], estas técnicas geralmente combinam levantamento de distribuições de frequências e detecção de anomalias. Um exemplo é a aplicação da Lei de Benford para detecção de fraudes contábeis, em que se assume uma determinada distribuição assintótica para o primeiro dígito significativo de dados obtidos através de distribuições variadas. A força da técnica está baseada na dificuldade em se gerar, a partir de dados fabricados, uma distribuição que esteja em conformidade com esta lei.

Através de uma revisão das técnicas utilizadas para detecção de fraudes entre 2004 e 2015, Albashrawi [16] concluiu que as técnicas mais largamente utilizadas são Regressão Logística, Redes Neurais, Árvore de Decisão, Máquinas de Vetores Suporte (SVM), Naive Bayes e Redes Bayesianas. Estas técnicas são utilizadas por mais de 50% dos artigos pesquisados pelo autor. Além de detecção de fraudes, o autor ressalta a utilidade das técnicas pesquisadas em grande parte de aplicações de negócios, como previsão de vendas ou falência de empresas.

Dos quatro grupos de fraude analisados por Albashrawi [16] (fraudes em declarações financeiras, fraudes bancárias, fraudes em seguros e outras fraudes financeiras), o tipo de análise mais predominante diz respeito às fraudes em declarações financeiras. As fraudes baseadas em declarações financeiras falsas constituem um grupo amplo de fraudes que consiste em, de acordo com Spathis [17], manipular intencionalmente informações da organização de forma a supervalorizar ativos, vendas e lucros, bem como subestimar fragilidades, despesas ou perdas. A existência de fornecedores de fachada se encaixa melhor neste grupo, uma vez que, contratados pelo Governo Federal, esses fornecedores oferecem declarações enganosas a respeito de sua real capacidade em cumprir suas obrigações no fornecimento dos bens e serviços acordados.

No âmbito da CGU, embora não existam modelos preditivos específicos para detecção de fornecedores de fachada, podemos citar a utilização de um modelo de ranqueamento de contratos públicos por risco, desenvolvido por Sales [3], que utiliza, dentre suas abordagens, o risco dos fornecedores, estimado através de informações da capacidade operacional das empresas, como quantidade de funcionários, e do histórico de contratações (por exemplo, se a empresa já foi punida no passado). O risco de contratos, neste caso, é estimado

via Regressão Logística, e os contratos são ranqueados para priorização das ações de auditoria. O modelo foi posteriormente aperfeiçoado por Sun e Sales em [18] através da aplicação de Redes Neurais Profundas.

Como o modelo atual não utiliza informações de fontes não estruturadas, as informações de classes aprendidas a partir de cenários de sedes de fornecedores podem ser um aperfeiçoamento útil para este modelo de ranqueamento de contratos, além de poderem ser utilizadas em conjunto com algumas tipologias de risco elaboradas no âmbito da CGU, conforme citado no Capítulo 1.

2.2.2 Reconhecimento de Cenários

Nos primórdios da pesquisa em visão computacional, o reconhecimento de cenários era visto em geral como resultado da integração entre os objetos que pertenciam à imagem, como formas, bordas e ângulos específicos. Exemplos desses primeiros trabalhos estão nos detectores de linhas e ângulos de Griffith [19][20].

O primeiro modelo amplamente conhecido na literatura a propor uma representação holística de cenários em um ambiente computacional, ou seja, uma representação que vai além de meros conjuntos de objetos, foi proposto por Oliva e Torralba [21]. O modelo, conhecido como “Envelope Espacial”, tenta capturar as propriedades globais de um cenário (ex. se o cenário é artificial ou natural ou é um ambiente aberto ou fechado).

No final da década passada, o reconhecimento de cenários utilizando técnicas baseadas em levantamento de atributos tornou-se bastante popular, como notado por Zou et al. em [22]. Uma das principais tentativas nesta direção foi proposta por Fei-Fei e Perona [23], que decompuseram conjuntos de imagens em uma série de atributos denominados *codewords*. O objetivo da fase de treinamento deste modelo era construir, a partir de grupos de imagens de entrada, uma determinada distribuição de *codewords* para cada grupo. Na fase de teste, uma imagem particular é decomposta em *codewords* e então comparada às distribuições construídas durante a fase de treinamento.

O modelo *Bag-of-Visual-Words* (BoVW), avaliado por Yang et al. [24], também foi uma técnica de reconhecimento de cenários muito popular na década passada. Esta técnica, também baseada em levantamento de atributos, é similar aos modelos de classificação de textos, porém utilizando grupos de pontos de interesse na imagem em vez de palavras. Cada ponto de interesse é tratado como uma “palavra visual” do conjunto de palavras que representam a imagem. Este conjunto é representado como um vetor contendo a frequência de cada palavra visual na imagem (similar à matriz “termo-documento” muito utilizada em classificação de textos). Este vetor pode então ser processado como um conjunto de atributos que, através de métricas de distância associadas, pode ser classificado em grupos similares. Uma desvantagem deste modelo é não representar o aspecto espa-

cial da imagem, pois tudo o que sabemos são atributos e suas frequências, mas não suas localizações, o que acaba gerando perda de informação relevante para a classificação.

Uma melhoria do modelo BoVW foi o *Spatial Pyramid Matching Kernel* (SPMK), proposto por Lazebnik, Schmid e Ponce [25]. Da mesma forma que o modelo BoVW, o SPMK representa a imagem como um conjunto de atributos. Porém, além de gerar este conjunto para imagem completa, gera também para subdivisões da imagem, de modo que o aspecto espacial da imagem seja melhor representado.

A partir de 2012, o estado da arte em visão computacional sofreu uma mudança significativa com o modelo AlexNet de Krizhevsky, Sutskever e Hinton [26], quando as redes neurais convolucionais (CNN) foram aplicadas com bastante sucesso nos problemas de classificação de imagens propostos pela competição Imagenet ILSVRC-2012. O modelo venceu a competição com uma taxa de erro de 15,3%, em contraste com a taxa de 26,2% obtida pelo segundo colocado² (Harada e Kuniyoshi [28]). Desde então, os modelos de estado da arte em reconhecimento de objetos e cenários vêm utilizando as CNN.

2.2.3 Arquiteturas ResNet

Com a popularização das redes convolucionais a partir de 2012, foram criadas arquiteturas de redes neurais cada vez mais profundas, conseqüentemente com capacidade de reconhecer padrões cada vez mais complexos. Começando pela AlexNet (de oito camadas), surgiram redes de 16 (VGG16) e 22 camadas (GoogLeNet).

Entretanto, o crescimento das redes além desses limites levava a problemas de degradação de desempenho nos modelos. De acordo com He et al. [29], estes problemas não são causados por *overfitting*, pois é possível verificar a degradação de desempenho em redes mais profundas no próprio conjunto de treinamento (ver erro da rede mais profunda - linha vermelha - versus erro da rede menos profunda - linha amarela da Figura 2.7).

Para combater esse problema, He et al. [29] propuseram uma solução por construção pelo seguinte princípio: dada uma rede neural com camadas copiadas de uma rede menos profunda, ao adicionarmos camadas que computam a função identidade ($f(x) = x$) em qualquer parte da rede, o modelo mais profundo construído deve possuir um erro de treinamento não superior ao da rede menos profunda. Na prática contudo, o processo de construção de um bloco de camadas convolucionais que se aproximasse de uma função identidade era computacionalmente complexo.

A solução encontrada por He et al. [29] foi construir blocos de camadas (chamados de blocos residuais) organizados da forma mostrada na Figura 2.8.

²Conforme Russakovsky et al. [27], a taxa de erro foi calculada utilizando a porcentagem de erros de classificação cometidos em relação às cinco categorias de objetos mais prováveis retornadas pelo modelo. Matematicamente, temos $erro = \sum_{i=1, j=1}^{i=N, j=5} \min_j d_{ij}$.

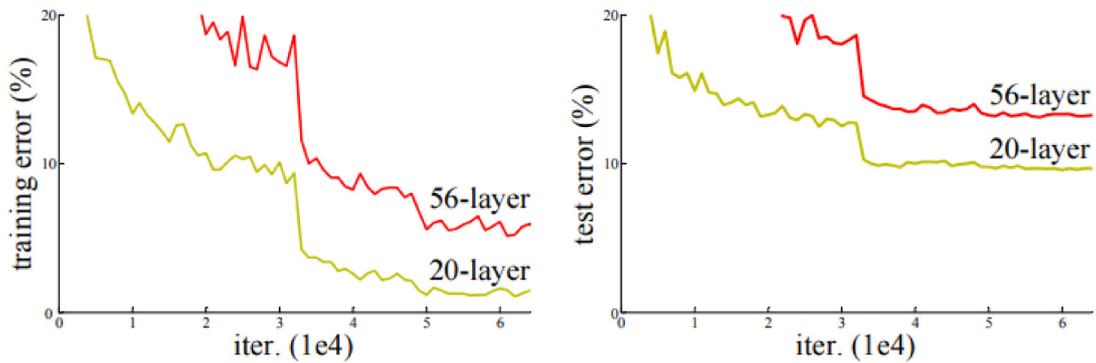


Figura 2.7: Exemplo de degradação no desempenho em redes com mais camadas (linha vermelha).

Fonte: He et al. (2015) [29].

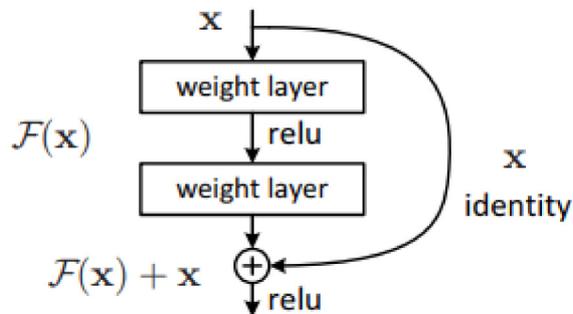


Figura 2.8: Exemplo de um bloco residual típico de uma arquitetura ResNet.

Fonte: He et al. (2015) [29].

Matematicamente, um bloco de camadas que computa a função $H(x)$ é transformado no bloco residual que computa a função $H(x) = F(x) + x$, sendo que a função $F(x) = H(x) - x$ é chamada de função residual.

Embora a construção de um bloco de camadas que compute a função identidade seja complexa computacionalmente, construir um bloco de camadas de modo a fazer a função residual se aproximar de zero ($F(x) \approx 0$) é significativamente mais simples. Desta forma, o bloco é construído de forma a minimizar esta função para aproximar este bloco da função identidade ($H(x) \approx x$, quando $F(x)$ tende a zero).

Isto permitiu a construção de redes convolucionais muito mais profundas ao mesmo tempo em que se evitou a degradação em seu desempenho. Existem redes de arquitetura ResNet com mais de uma centena de camadas (tais como a ResNet152 [29]).

Entretanto, os blocos residuais acabam por compartilhar muito pouca informação entre si. Isso também pode acabar por limitar o desempenho da rede, o que é conhecido como

problema da diminuição do reuso (*diminishing feature reuse*).

Para combater o problema, foi criada uma variação da arquitetura ResNet denominada Wide ResNet [30], cujas redes possuem camadas mais largas (com mais neurônios) e menos profundas que as ResNets tradicionais (ver Figura 2.9). Como aumentam a taxa de reuso entre as camadas da rede e se aproveitam das capacidades de processamento paralelo das GPUs modernas, as Wide ResNets provaram ser mais velozes no treinamento, ao mesmo tempo em que possuem acurácia similar ou superior às demais ResNets.

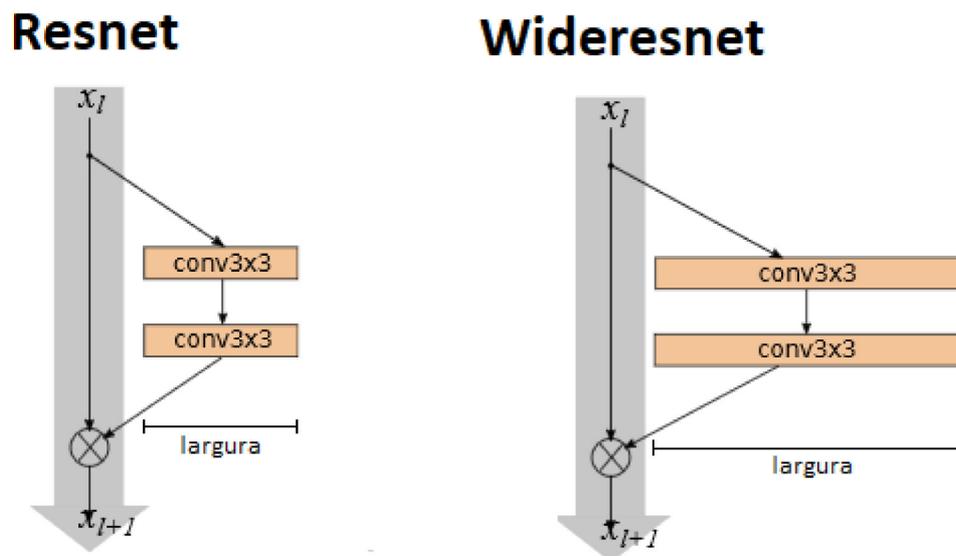


Figura 2.9: Exemplo de um bloco residual simples de uma ResNet e uma Wide ResNet.

Fonte: Zagoruyko e Komodakis (2016) [30].

O modelo que utilizamos é uma Wide ResNet de 18 camadas, pré-treinada no conjunto de cenários Places365, obtido de Zhou et al. [31].

A estrutura detalhada desta rede está na Figura 2.10. Nesta Figura, as operações de Convolução e *Pooling* são denotadas respectivamente por “conv” e “pool”. Os números que acompanham as operações de convolução (ex. 7x7 e 3x3) indicam o tamanho da janela que a operação de convolução utiliza para dividir a imagem de entrada.

A quantidade de filtros (*filters*) em cada camada convolucional equivale à quantidade de regiões de unidades locais (*feature maps*) em cada camada que compartilham os mesmos pesos. A operação de *Batch Normalization* é denotada por “norm” e “relu” é a função de ativação aplicada em certas camadas desta rede. Temos também algumas camadas que possuem função de subamostragem (*downsampling*), denotadas por “down”.

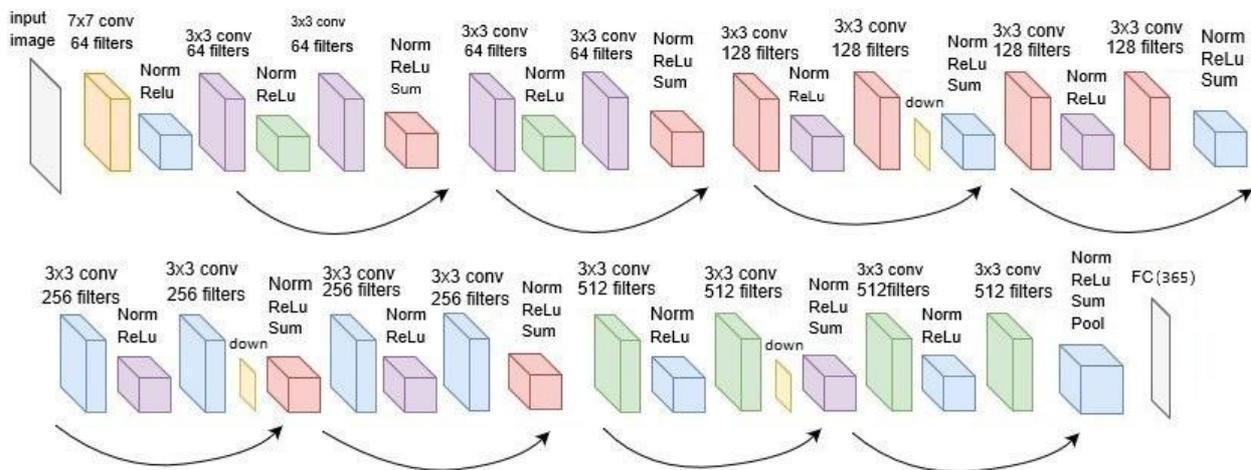


Figura 2.10: Arquitetura da Wide Resnet 18.

Fonte: Al Rabbani Alif, Ahmed e Hasan [32].

As setas indicam as funções identidade aplicadas nos oito blocos residuais existentes nesta rede. Na camada ao final de cada bloco residual, podemos ver também a operação de soma da função residual com a função identidade ($F(x) + x$), denotada por “sum”.

Cada bloco residual possui duas convoluções, totalizando 16 camadas convolucionais. Se adicionarmos a isto a camada convolucional inicial (7x7 conv) e a camada completamente conectada ao final da rede (denotada por “FC”), temos no total 18 camadas principais para esta rede, fazendo referência ao nome desta arquitetura.

2.2.4 Bancos de Dados de Cenários

O avanço das técnicas que utilizam redes neurais profundas em tarefas como detecção de padrões, visão computacional e processamento de linguagem natural deve-se não somente ao avanço de arquiteturas e algoritmos, mas também à disponibilidade de grandes quantidades de dados. De acordo com Zhou et al. [31], os avanços de softwares de Inteligência Artificial em jogos, como o caso do *Deep Blue* no Xadrez, do Watson no jogo de perguntas e respostas *Jeopardy* e, mais recentemente, do *AlphaGo* em Go, veio acompanhado de grandes bancos de dados, com 700.000, 8,6 milhões e 30 milhões de itens, respectivamente. Entretanto, o maior exemplo de avanço técnico baseado em grandes quantidades de dados vem, entretanto, do banco de dados do *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC). O banco de dados desta competição, que acontece desde 2010 e é considerada por muitos os “jogos olímpicos” da Visão Computacional, possui milhões de imagens classificadas entre centenas de categorias de objetos (Russakovsky et al. [33]).

A rede ResNet vencedora desta competição em 2015 (He et al. [29]) obteve uma taxa de erros em classificação de objetos menor que 4%, superando a performance humana, que possui taxa de erros em classificação de objetos próxima de 5%.

Analogamente aos casos mencionados, também a área de reconhecimento de cenários depende da evolução de conjuntos de dados orientados a cenários com coberturas cada vez maiores e mais diversas, além de mais densas (isto é, com maior quantidade de exemplos por categoria). Alguns bancos de dados de cenários criados ao longo dos anos são SUN397 [34], MIT Indoor67 [35], Scene15 [25] e SUN Attribute [36]. Entretanto, o conjunto de dados de cenários com maior cobertura e densidade atualmente é o banco Places. Criado por Zhou et al. [31] a partir do banco SUN397, este banco é composto por mais de 10 milhões de imagens divididas entre 434 categorias de cenário.

Como ponto de partida deste projeto, utilizamos os atributos aprendidos de uma rede de arquitetura Wide ResNet pré-treinada em um subconjunto do banco Places denominado Places365. De acordo com Zhou et al. [31], o conjunto Places365 foi obtido através da seleção de 365 categorias de cenários com mais de 4.000 imagens cada (entre as 434 classes originais do banco Places). As imagens se dividem em três grandes categorias: Internas (como quartos, cafeterias, bares), Externas Naturais (estradas, florestas) e Externas Urbanas (torres, arcos, estações de trem). Possui 1.803.460 imagens de treinamento, com número de imagens por classe variando de 3.068 a 5.000. O conjunto de validação possui 50 imagens por classe, e o de teste, 900 imagens por classe.

Capítulo 3

Solução Proposta

Neste capítulo, abordamos as principais etapas a serem executadas ao longo do trabalho para atingirmos os objetivos geral e específicos apresentados na Seção 1.4. Algumas medidas de sucesso que poderão ser utilizadas para verificar o atendimento destes objetivos a longo prazo são: aumento na quantidade de fornecedores de fachada ou de alto risco descobertos pela CGU e suas unidades regionais durante o período de um ano com o auxílio do trabalho; e quantidade de esquemas de corrupção e desvios de recursos públicos investigados como decorrência da descoberta de fornecedores de fachada através da implantação deste trabalho.

Para que o projeto seja executado, além das imagens das sedes de fornecedores, é necessário o acesso a algumas bases de dados do Governo Federal, listadas a seguir:

- base do Sistema de Administração de Serviços Gerais (Comprasnet SIASG): Contém informações a respeito das licitações e contratos do Poder Executivo Federal, bem como de órgãos de outros poderes e esferas governamentais. Através desta base é possível obter informações sobre contratos firmados com o Governo Federal e seus fornecedores. Será útil ao trabalho não somente por permitir a obtenção de fornecedores e contratos, mas também por permitir inferências sobre a capacidade financeira das empresas contratadas (obtidas através de dados de contratos anteriores). A base do Comprasnet SIASG está disponível em grande parte de forma aberta através da API de Compras Governamentais (<http://compras.dados.gov.br/docs/home.html>);
- base de Pessoas Físicas e Jurídicas da Receita Federal do Brasil (RFB): Contém os endereços dos fornecedores que serão utilizados como base para as consultas realizadas na API do Google Street View. Além disso, permite a obtenção de informações de sócios e suas respectivas qualificações, que podem ser úteis na construção de um

modelo mais geral de aprendizagem de máquina para detecção de fornecedores de fachada;

- base da Relação Anual de Informações Sociais (RAIS): Contém informações sócio-econômicas de empregados e empresas, solicitadas pelo Ministério do Trabalho e Emprego. A partir desses dados, é possível determinar informações importantes para a caracterização de fornecedores de fachada, como a quantidade e o nível de escolaridade de empregados de uma organização;
- base da Classificação Nacional de Atividades Econômicas (CNAE): pertencente à RFB, contém descrições e códigos de atividade econômica de todos os agentes econômicos engajados na produção de bens e serviços, incluindo estabelecimentos públicos e privados, além de instituições sem fins lucrativos e agentes autônomos. Pode ser utilizada em conjunto com os dados do Comprasnet SIASG a fim de verificar se os fornecedores estão oferecendo bens e serviços compatíveis com a atividade econômica proveniente deste cadastro;
- base do Cadastro de Entidades Inidôneas e Suspensas (CEIS): é um cadastro próprio da CGU, que lista empresas que foram declaradas inidôneas pela Administração Pública e estão suspensas de contratar com a Administração até o fim do prazo da sanção, que varia de acordo com a gravidade da falta cometida pelo fornecedor. Este cadastro nos ajuda a levantar tipologias de risco que podem estar relacionadas a fornecedores de fachada.

Para atingirmos os objetivos definidos, além das metas de longo prazo, utilizamos critérios específicos para as atividades de Mineração de Dados. Para medir o desempenho do classificador de imagens, usamos a Área sobre a Curva ROC; para medir o desempenho da classificação de fornecedores de risco em geral, utilizamos as medidas F-Measure e a Área sobre a Curva ROC; Por fim, é necessário incluir uma medida da importância do classificador de cenários no modelo de risco mais geral, dado que uma das hipóteses é verificar se a inclusão dessa informação melhora o desempenho de um classificador mais geral de risco. Mais uma vez, medimos essa melhora com base nas medidas F-Measure e Área sobre a Curva ROC, comparando as medidas entre um modelo que não utiliza informações do classificador de imagens com um modelo que utiliza essas informações.

3.1 Metodologia

Para que possamos atingir as metas estipuladas ao longo deste projeto, propomos como abordagem para sua execução o seguinte fluxo de tarefas:

1. obter as imagens dos fornecedores;
2. realizar anotações nas imagens obtidas;
3. ajustar modelos baseados em redes convolucionais;
4. construir modelos de risco gerais com diversos atributos, utilizando a saída do modelo convolucional;
5. realizar a validação dos modelos de risco gerais, avaliando se houve melhoria pela inclusão da saída do modelo convolucional.

Explicamos cada uma das tarefas propostas nas próximas seções. A metodologia para atividades de mineração de dados foi em parte baseada na metodologia do padrão *Cross Industry Standard Process for Data Mining* (CRISP-DM) [37].

3.1.1 Obter as imagens dos fornecedores

A API do Google Street View nos fornece basicamente duas formas de obtenção de imagens: através de coordenadas específicas ou através de endereços. Infelizmente, não há atualmente nas bases de dados do Governo Federal informações a respeito das coordenadas específicas das sedes dos fornecedores. Entretanto, ao cruzarmos dados do Comprasnet SIASG com dados da RFB, é possível obter os endereços dos fornecedores com contratos ativos com o Poder Executivo Federal, o que, na maior parte dos casos¹, é o suficiente para acessarmos as imagens de suas sedes através do Google Street View.

Entretanto, há ainda casos em que os endereços obtidos não estão cobertos pelo Google Street View. Nestes casos, é necessário verificar se há imagens disponíveis para o local pesquisado, o que pode ser feito através de consultas aos metadados da imagem pela API do Street View (ver Figura 3.1). Após a requisição pelos metadados, a API retorna status de resposta “OK” caso seja possível recuperar a imagem com sucesso².

Para minimizarmos erros relacionados ao download de imagens em localizações incorretas, todos os endereços utilizados como entrada para a API foram formatados da seguinte maneira: “(Tipo de Logradouro). (Logradouro), (Número) - (Bairro), (Cidade) - (Sigla da UF), (CEP)”, onde os campos entre parênteses são campos presentes na tabela de pessoa jurídica da Receita Federal. O campo “Complemento” do endereço foi excluído desta formatação por verificarmos que sua inclusão muitas vezes confunde a API do Street View, levando a endereços incorretos. A Figura 3.1 mostra um exemplo de requisição por imagem utilizando o endereço formatado, bem como a respectiva imagem de resposta.

¹Em alguns casos, os endereços podem estar incompletos ou incorretos, levando a localizações erradas ou não existentes.

²Documentação disponível para consulta em <https://developers.google.com/maps/documentation/streetview/metadata?hl=pt>.

Requisições GET

Metadados:

`https://maps.googleapis.com/maps/api/streetview/metadata?size=227x227&location=R. ALFREDO WAGNER, 858 - ALVORADA, CHAPECO-SC,89804-430 &heading=90&key=CHAVE`



Respostas

```
{
  "copyright" : "© Google, Inc.",
  "date" : "2012-04",
  "location" : {
    "lat" : -27.07809942823494,
    "lng" : -52.63945563168106
  },
  "pano_id" : "BfnAWN0g79DwgTS3lqBt0g",
  "status" : "OK"
}
```

Imagens:

`https://maps.googleapis.com/maps/api/streetview?size=227x227&location=R. ALFREDO WAGNER, 858 - ALVORADA, CHAPECO-SC,89804-430&heading=90 &key=CHAVE`

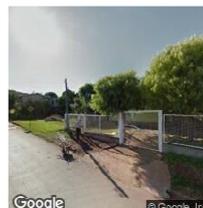


Figura 3.1: Exemplo de requisições HTTP GET para a API do Street View para a obtenção de metadados e imagens.

Cabe ressaltar neste ponto a opção de obter imagens através da API do Street View e não da API do Google Maps. Enquanto o Google Maps apresenta imagens de satélite, o Google Street View apresenta imagens em nível de rua. Consideramos porém que a utilização de imagens em nível de rua poderia nos fornecer uma quantidade muito maior de informações de cenários (em termos de quantidade e variedade de classes) que a utilização de imagens de satélite.

Além disso, a opção pela utilização do Street View também está ligada à opção de utilizarmos inicialmente uma rede pré-treinada a partir do conjunto de cenários Places365 [31], que, a exemplo do Street View, é composto apenas de imagens em nível de rua.

Entretanto, a classificação de imagens de satélite do Google Maps pode servir de bom complemento à classificação de imagens obtidas pelo Street View, podendo ser incluídas em trabalhos futuros através do treinamento de uma rede convolucional própria para este tipo de imagem.

Dos 37.103 fornecedores encontrados na base do ComprasNet com contratos ativos entre janeiro de 2016 e dezembro de 2017 com o Poder Executivo Federal, foi possível baixar imagens de 32.482 fornecedores via Google Street View. O período escolhido para os contratos levou em conta a necessidade de atuação tempestiva dos auditores da CGU em possíveis casos de fornecedores de fachada, bem como de delimitação de escopo para a conclusão deste trabalho. Contratos de outros períodos podem naturalmente ser adicionados no futuro.

Para cada fornecedor, foram baixadas 4 imagens, orientadas em relação aos pontos

cardeais³, totalizando portanto 129.928 imagens baixadas. Entendemos ser necessário capturar essas quatro perspectivas pelo fato de apenas uma imagem por fornecedor muitas vezes não ser suficiente para capturar as características de um cenário (muitas imagens apresentam apenas ruas e paredes, por exemplo).

As imagens foram baixadas em tamanho 227 x 227 em 3 canais (RGB), inicialmente utilizadas como entrada para uma rede convolucional mais simples (Alexnet). Ao utilizarmos a arquitetura Wide ResNet foi necessário redimensionar as imagens baixadas para o tamanho 224 x 224. Em uma implementação utilizando a linguagem Python e quatro *threads* em um computador simples de quatro núcleos, foram necessárias aproximadamente doze horas para baixar quase 130 mil imagens.

Das imagens baixadas, entretanto, nem todas são de origem da própria Google. Há imagens de colaboradores individuais, cobrindo cenários que vão além dos de rua, como ambientes internos e naturais (como praias, parques e cachoeiras). Por diferirem bastante das demais imagens, decidimos por excluí-las e manter apenas as imagens de origem da Google (a autoria das imagens está presente nos metadados, no campo “copyright”). Além deste filtro, consideramos apenas imagens de fornecedores que possuíam atividade econômica registrada na base da CNAE, para facilitar cruzamentos futuros com essa base.

Após a aplicação desses filtros, obtivemos um conjunto final de 120.752 imagens relativas a 30.188 fornecedores.

3.1.2 Realizar anotações nas imagens obtidas

Após a obtenção das imagens, anotamos exemplos positivos e negativos de cenários comerciais através de informação visual. Cenários positivos possuem características como letreiros com nomes de lojas e estacionamentos, enquanto as imagens negativas não apresentam nenhuma destas características (como uma residência ou terreno baldio).

Seguindo esta metodologia, anotamos exemplos de 4.116 cenários comerciais (positivos) e 4.128 cenários não comerciais (negativos), todos eles extraídos de fornecedores do Governo Federal da etapa anterior, conforme imagens da Figura 3.2 de nossa pesquisa em Wacker, Ferreira e Ladeira [38].

³Orientação dada pelo parâmetro “heading” da API do Street View, sendo “heading=0”, ou 0 grau, indicando a direção norte. As direções leste, sul e oeste são indicadas, respectivamente, pelos valores 90, 180 e 270 graus.



Figura 3.2: Exemplo de imagens anotadas como áreas comerciais e não comerciais.

Fonte: Wacker, Ferreira e Ladeira (2018) [38].

3.1.3 Ajustar modelos baseados em redes convolucionais

Um dos principais problemas encontrados ao utilizarmos uma rede pré-treinada sem nenhum ajuste é o fato de que as classificações encontradas podem não ser adequadas ao nosso contexto.

A Figura 3.3 ilustra o fato. Ao utilizarmos uma rede convolucional simples (AlexNet) treinada a partir do banco de dados Places205 [39], precursor do banco Places365 [31], é possível ver que a rede acerta categorias mais comuns (imagens à esquerda na Figura). Entretanto, ao nos aproximarmos de cenários residenciais mais próximos da realidade brasileira (imagens à direita), ocorrem erros de classificação.



Figura 3.3: Exemplo de imagens classificadas pela rede convolucional AlexNet treinada no conjunto de dados Places205.

Adicionalmente, se observarmos a Figura 3.4, extraída de nossa pesquisa em Wacker, Ferreira e Ladeira [38], podemos ver como pode ser útil utilizar modelos pré-treinados. A Figura mostra imagens de fornecedores marcadas pelas regiões que mais ativaram os neurônios da penúltima camada da rede, onde cada linha representa ativações de um tipo de neurônio. O primeiro tipo de neurônio parece identificar carros, o segundo complexos habitacionais e o último ambientes comerciais. Isso nos mostra que a rede pré-treinada já é capaz de reconhecer atributos úteis para classificação de cenários, que podem ser aproveitados via técnicas de *Transfer Learning*.

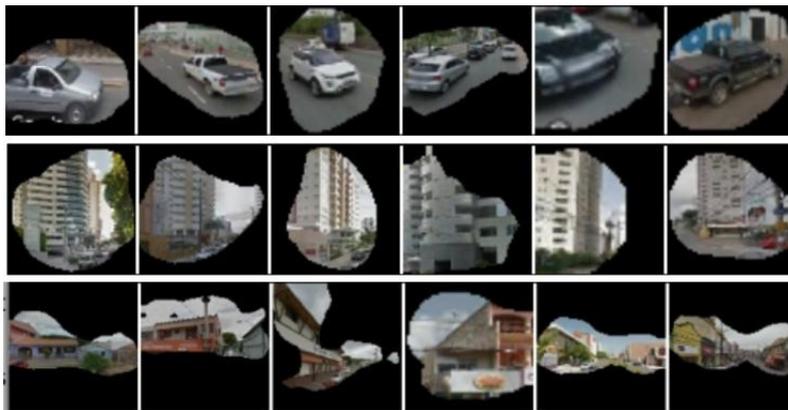


Figura 3.4: Exemplo de regiões que mais ativaram neurônios da penúltima camada da rede Wide ResNet18.

Fonte: Wacker, Ferreira e Ladeira (2018) [38].

Neste trabalho, foram utilizadas as duas abordagens mostradas na Seção 2.1.6 (extração de atributos e ajuste fino) para realizar *Transfer Learning* na arquitetura WideResNet treinada a partir do banco de dados Places365. Os resultados e métricas de desempenho dessas duas técnicas estão na análise de resultados do próximo capítulo.

3.1.4 Construir modelos de risco gerais com diversos atributos

Nesta etapa, utilizamos a saída do modelo convolucional com melhor performance da etapa anterior em conjunto com tipologias levantadas no Observatório da Despesa Pública da CGU (ODP), a fim de avaliar se as informações extraídas da rede convolucional serão úteis como indicador adicional de risco de fornecedores.

Para isso, construímos modelos básicos de Regressão Logística, que utilizarão, além da saída do modelo convolucional (probabilidades de um cenário ser comercial ou não), dez tipologias levantadas pelo ODP:

- tipologia 1 - Baixa quantidade de funcionários: esta tipologia identifica fornecedores que tiveram uma quantidade média de funcionários inferior a 10 nos últimos dois

anos. Utiliza como fontes de dados a base do Sistema de Compras Governamentais (Comprasnet SIASG) para identificação dos fornecedores, bem como as bases da Relação Anual de Informações Sociais (RAIS) de 2016 e 2017 para cálculo da quantidade média de funcionários nos últimos dois anos;

- tipologia 2 - Sócios com baixa qualificação ou remuneração: identifica sócios de fornecedores com vínculo empregatício de baixa remuneração (inferior a 2 salários mínimos) ou qualificação (constante de uma lista de códigos da classificação brasileira de ocupações utilizada pelos auditores da CGU). Utiliza como fonte de dados a base do Comprasnet SIASG para fornecedores, da Receita Federal do Brasil (RFB) para extração de vínculos societários e a base da RAIS de 2017 para vínculos empregatícios;
- tipologia 3 - Microempresas ou empresas de pequeno porte com licitações de valores incompatíveis: busca microempresas ou empresas de pequeno porte que venceram licitações cujo valor total em 2017 supera os limites legais para que elas sejam ainda consideradas microempresas ou empresas de pequeno porte. Pela Lei Complementar 123/2006, que regula o tema, este limite é de R\$ 360.000,00 para microempresas e R\$ 3.600.000,00 para empresas de pequeno porte. Utiliza a base do Comprasnet SIASG para cálculo dos valores totais vencidos pelas empresas, bem como para obtenção de seu porte;
- tipologia 4 - Mais de um fornecedor no mesmo endereço: lista fornecedores que possuem endereços com descrições idênticas. Utiliza as bases Comprasnet SIASG e RFB;
- tipologia 5 - Grande quantidade de subclasses de atividade econômica secundária: busca fornecedores com mais de 10 subclasses de atividade econômica secundária. Utiliza como fontes a base do Comprasnet SIASG e a base da Classificação Nacional de Atividades Econômicas (CNAE) da Receita Federal, a fim de obter todas as atividades econômicas de um fornecedor;
- tipologia 6 - Diferença entre data da abertura e compra inferior a 30 dias: lista fornecedores que ganharam licitações a menos de 30 dias da data de abertura da empresa. Utiliza como fontes a base do Comprasnet SIASG e a base da RFB para encontrar a data de abertura do fornecedor;
- tipologia 7 - Fornecedores com penalidades suspensivas no cadastro: Fornecedores que foram punidos com penalidades graves (suspensivas), que impedem a contratação com a Administração Pública por determinado período. As penalidades con-

sideradas são suspensão, declaração de inidoneidade e impedimento do fornecedor. Utiliza como fonte a base do Comprasnet SIASG;

- tipologia 8 - Empresas no CEIS (Cadastro de Entidades Inidôneas e Suspensas): Fornecedores que constam ou constaram do Cadastro de Entidades Inidôneas e Suspensas da CGU. Utiliza como fontes a base do Comprasnet SIASG e a base do Cadastro de Entidades Inidôneas e Suspensas (CEIS) da CGU;
- tipologia 9 - Empresas com sócios com empresas no CEIS: Fornecedores que não possuem cadastro no CEIS, mas com sócios possuidores de outras empresas que constam ou constaram desse cadastro. Utiliza como fontes a base do Comprasnet SIASG, da RFB e do CEIS da CGU;
- tipologia 10 - Compatibilidade entre atividade econômica e produto/serviço fornecido: verifica se o tipo de produto ou serviço fornecido é compatível com as classificações de atividade econômica do fornecedor. Utiliza a base do Comprasnet SIASG para obter fornecedores e tipos produtos e serviços fornecidos, a base CNAE da Receita Federal para obter as atividades econômicas dos fornecedores e uma tabela de correspondências válidas entre código de material ou serviço e atividade econômica.

Tipicamente os auditores da CGU utilizam estas tipologias como um conjunto de indicadores que, totalizados de forma simples, podem direcionar ações de auditoria envolvendo determinados fornecedores e contratos.

No Capítulo 4, apresentamos uma comparação da saída do modelo convolucional com as tipologias aqui apresentadas, bem como os Modelos de Regressão Logística que utilizam essas tipologias em conjunto com a saída da rede convolucional, tomando como variável dependente a tipologia 7 (Fornecedores com penalidades suspensivas).

3.1.5 Realizar a validação dos modelos de risco gerais

Na etapa final, fazemos uma comparação entre diversos ajustes de modelos de risco com Regressão Logística utilizando as tipologias do ODP, tomando sempre como variável-alvo a tipologia 7 (Fornecedores com penalidades suspensivas). Os modelos comparados são:

1. modelo de regressão logística com base em variáveis binárias das tipologias levantadas pelo ODP;
2. modelo de regressão logística com base nas variáveis quantitativas das tipologias levantadas pelo ODP;
3. modelos de regressão logística separados por seção da Classificação Nacional de Atividades Econômicas (CNAE).

Os resultados da comparação entre esses modelos, bem como da verificação da relevância da variável extraída da rede convolucional, podem ser vistos no próximo capítulo.

Capítulo 4

Resultados

Neste capítulo apresentamos os resultados de comparação das técnicas de transferência de aprendizado da rede convolucional pré-treinada (Wide ResNet 18) para tornar o classificador capaz de diferenciar entre cenários comerciais e não comerciais. Fizemos também testes estatísticos comparando esta classificação com as demais tipologias levantadas pelo Observatório da Despesa Pública. Por fim, unificamos estas tipologias em alguns modelos de risco gerais com base em fornecedores punidos utilizando Regressão Logística, a fim de avaliar se a classificação dada pela rede convolucional é uma informação capaz de melhorar a performance preditiva desses modelos.

4.1 *Transfer Learning* da Rede Convolucional (CNN)

Conforme metodologia explicada em nossa pesquisa em Wacker, Ferreira e Ladeira [38], utilizando a rede WideResNet de 18 camadas pré-treinada no conjunto Places365, foram extraídos atributos de saída das seis últimas camadas de convolução deste modelo (que chamamos de Conv-12 a Conv-17). A esses vetores de atributos foram aplicadas operações de *pooling* para redução de dimensionalidade, de forma que todos os atributos extraídos fossem transformados em vetores de 512 dimensões (com exceção da camada Conv-12, cuja saída é um vetor de 256 dimensões). Foram treinados modelos SVM com os atributos extraídos de Conv-12 a Conv-17, tendo como classe-alvo as categorias de cenários comerciais e não comerciais anotadas manualmente.

Estes modelos foram comparados com a abordagem de ajuste fino (*Fine-Tuning*) da rede, onde construímos uma rede convolucional adaptada através da substituição da última camada (com saída correspondente às 365 categorias de cenário do modelo original) para uma camada com apenas duas saídas, correspondentes às novas categorias de cenários comerciais e não comerciais que foram anotadas manualmente. A partir daí, continuamos o treinamento desta nova rede (ajustada com dois neurônios na camada de saída), porém

mantendo os pesos das primeiras 12 camadas inalterados (ajustando portanto apenas os parâmetros das últimas 6 camadas)¹.

A comparação de desempenho entre as abordagens propostas, utilizando como métrica a área sobre a curva ROC (AUC), está nas Tabelas 4.1 e 4.2, extraídas de nossa pesquisa em Wacker, Ferreira e Ladeira [38].

Tabela 4.1: Classificação de imagens individuais: Média \pm Desvio Padrão AUC sobre 10 rodadas (com exceção do fine-tuning) para vários atributos.

Atributos extraídos da Places	Média da AUC
Conv-12	0.8877 \pm 0.0136
Conv-13	0.9208 \pm 0.0078
Conv-14	0.9387 \pm 0.0067
Conv-15	0.8527 \pm 0.0129
Conv-16	0.9545 \pm 0.0056
Conv-17	0.9483 \pm 0.0073

Tabela 4.2: Área sobre a curva ROC (AUC) da abordagem de Fine-Tuning da Wide Resnet18.

Fine-Tuning (AUC)
0.9737

Como podemos verificar, o modelo de melhor desempenho apresentado foi o construído através do ajuste fino das últimas 6 camadas da rede, com área sobre a curva ROC igual a 0,9737. Dentre os modelos de *Transfer Learning* utilizando o SVM com atributos extraídos da rede, os melhores modelos utilizaram atributos das últimas camadas convolucionais (Conv-16 e Conv-17), o que está de acordo com a intuição de que os atributos mais complexos e representativos estão concentrados nas últimas camadas da rede. Nossa pesquisa em Wacker, Ferreira e Ladeira [38] mostrou ainda que o desempenho dos modelos não é influenciado pelos ângulos de orientação das imagens baixadas. Ao considerarmos modelos treinados com imagens sempre orientadas a determinados ângulos (0, 90, 180 ou 270 graus), os resultados no desempenho são semelhantes.

¹Foi utilizado o algoritmo de otimização Adam com os seguintes hiperparâmetros: taxa de aprendizado de 0,0001, decaimento de pesos de 0,001, tamanho do lote de treinamento = 64 e quantidade de loops de treinamento (épocas) = 5.

4.2 Comparando a saída da CNN com tipologias do ODP

Utilizando o modelo de rede convolucional mais performático dentre os testados (aquele obtido com ajuste fino), executamos a predição de cenários comerciais ou não comerciais (com base nas imagens previamente anotadas) para as 120.752 imagens de 30.188 fornecedores obtidas durante a primeira etapa do trabalho.

Para cada uma das imagens, o modelo traz a probabilidade de ela representar um endereço comercial. Como cada fornecedor possui quatro imagens, respectivas às quatro orientações disponíveis, optamos por ser conservadores na análise e considerar um fornecedor em cenário comercial se pelo menos uma das imagens apresentar probabilidade maior que 50% de ser um cenário comercial, o que equivale a dizer que estamos extraíndo o máximo do vetor de quatro probabilidades disponíveis, conforme mostrado na Figura 4.1 adaptada de Wacker, Ferreira e Ladeira [38].

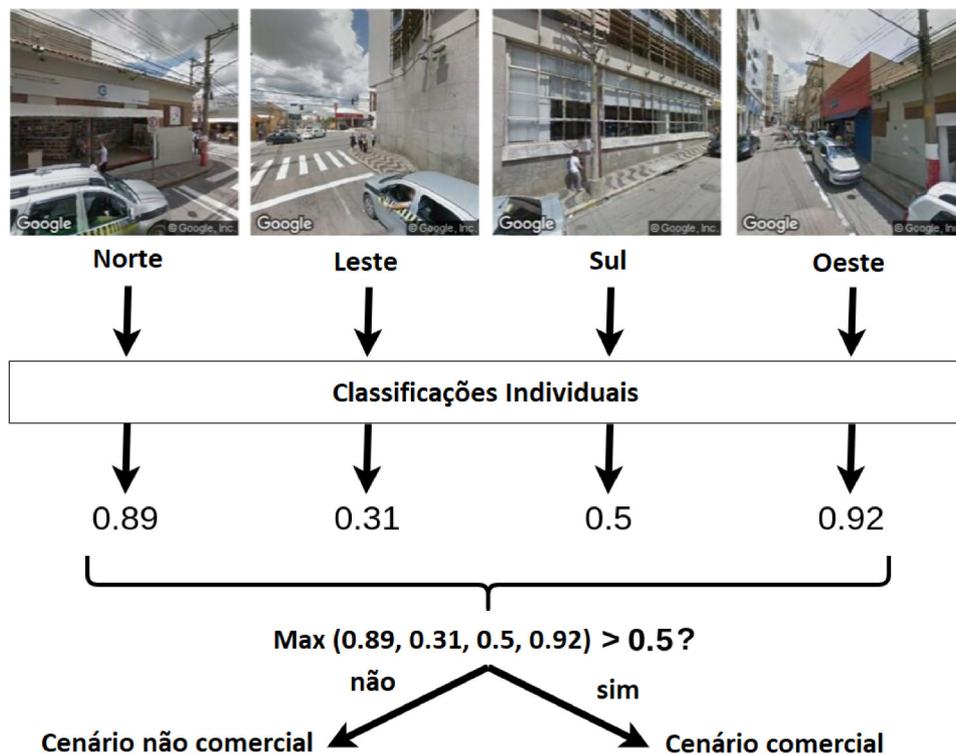


Figura 4.1: Abordagem utilizada para classificação dos cenários de fornecedores.

Fonte: Wacker, Ferreira e Ladeira (2018) [38].

A partir disso, construímos uma tabela com quatro campos: CNPJ do fornecedor (“CpfCnpjFornecedor”), probabilidade do fornecedor estar em um cenário comercial com

base no maior valor de probabilidade dentre as quatro imagens extraídas de cada fornecedor (“prob_loja”), e dois campos categóricos, um com valor booleano (“loja”) indicando se a probabilidade do fornecedor estar em um cenário comercial é superior a 50% e outro (“classificacao”) indicando se o fornecedor provavelmente está em cenário comercial (valor “Loja”, caso prob_loja $\geq 0,6$), indefinido (valor “Indefinido”, caso $0,4 < \text{prob_loja} < 0,6$) ou não comercial (valor “NaoLoja”, caso prob_loja $\leq 0,4$). As primeiras linhas dessa tabela estão ilustradas na Tabela 4.3.

Tabela 4.3: Campos derivados da saída da rede convolucional.

Cpf_Cnpj_Fornecedor	prob_loja	loja	classificacao
00000000034096	0.993297	True	Loja
00000000064327	0.999977	True	Loja
00000000105104	0.970694	True	Loja
00000000314706	0.926276	True	Loja
00000000429945	0.999390	True	Loja

Uma vez construída esta tabela, foram realizados diversos testes Qui-Quadrado a fim de verificar se o valor “Loja” do campo “classificacao” da tabela (ou seja, se o cenário do fornecedor for comercial) possui relação estatisticamente significativa² com as tipologias do ODP listadas no Capítulo 3, Seção 3.1.4. O resultado encontra-se na Tabela 4.4 (os números entre parênteses fazem referência ao número das tipologias listadas na Seção 3.1.4.).

O resultado encontrado na tabela mostra que cenários classificados como comerciais possuem maior frequência de casos de fornecedores com baixa quantidade de funcionários e esta relação possui significância estatística. Isso pode ocorrer devido à natureza das lojas anotadas ser essencialmente composta de lojas de rua (ver exemplos da Figura 3.2), que possuem menos funcionários se comparadas a lojas menos comuns nas imagens (como supermercados) ou mesmo a cenários classificados como não comerciais (por exemplo, uma universidade que possui muitos funcionários mas não se encaixa em um cenário comercial).

Observou-se também nestes relacionamentos que entre cenários comerciais há menos casos de microempresas ou empresas de pequeno porte que ganharam valores acima de seu limite em 2017, o que é consistente com a observação de que a maioria dos cenários de lojas anotados são lojas pequenas, portanto com menor possibilidade de ganhar licitações de valores grandes. Ainda seguindo a mesma lógica, lojas pequenas tendem a operar

²Para os testes estatísticos e análises de significância (p-valor) deste trabalho, consideramos o nível de significância $\alpha = 0.05$.

Tabela 4.4: Relacionamentos entre Cenário Comercial e outras Tipologias.

Tipologia	Efeito do relacionamento na frequência observada	P-valor do teste
Baixa quantidade de funcionários (1)	Maior que a esperada	0.007
Sócios com baixa qualificação ou remuneração (2)	-	0.47
EPP ou ME com valor homologado acima do limite (3)	Menor que o esperado	0.0015
Mais de um fornecedor com o mesmo endereço (4)	-	0.06
Grande quantidade de subclasses CNAE (5)	Menor que o esperado	7e-18
Fornecedor venceu licitações menos de 30 dias após abertura da empresa (6)	-	0.55
Fornecedor com punição suspensiva no cadastro (7)	Menor que o esperado	4e-14
Fornecedor cadastrado no CEIS (8)	-	0.0503
Fornecedor com sócios no CEIS (9)	Menor que o esperado	0.04
Compatibilidade entre atividade econômica e produto/serviço (10)	-	0.69

em menos atividades econômicas distintas, o que poderia explicar também mais este relacionamento.

Por fim, há uma relação significativa entre cenários classificados como comerciais e uma menor frequência de punições suspensivas no SICAF, bem como menor frequência de sócios com fornecedores no CEIS, o que sugere que fornecedores de risco se concentram mais no grupo de fornecedores com cenários classificados como não comerciais.

4.3 Avaliando modelos gerais de risco de fornecedores

Nesta seção, utilizamos modelos de Regressão Logística para analisar o relacionamento entre as variáveis de saída da rede convolucional (como “prob_loja” e “classificacao”) e as outras tipologias levantadas. A tipologia 7 (“Fornecedores com punições suspensivas no

cadastro”) foi utilizada como variável dependente em todos os modelos, enquanto que a tipologia 8 (“Fornecedores no CEIS”) foi excluída por apresentar alto grau de correlação com a tipologia 7 utilizada como variável dependente.

Todos os modelos utilizam a proporção de 2/3 dos dados para treinamento e 1/3 de dados para teste. Além disso, foi feito um *oversampling* nos dados de treinamento para equilibrar dados de fornecedores punidos com dados de fornecedores não punidos (no conjunto original, fornecedores punidos representam pouco menos de 1/3 do conjunto total analisado de 30 mil fornecedores). A proporção por *oversampling* transformou o conjunto de treino em um conjunto com 50% de dados de fornecedores punidos.

4.3.1 Modelo Base (Variáveis Binárias)

Primeiro criamos um modelo base utilizando Regressão Logística sem Parâmetros de Regularização com apenas variáveis binárias sobre as tipologias levantadas (0/1, True/False, relacionadas a se o fornecedor está abordado pela tipologia ou não), conforme ilustrado pela Tabela 4.5.

Tabela 4.5: Campos derivados das tipologias utilizadas para ajuste do modelo base.

Mais de 10 subclasses	EPP/ME acima do limite	ind_socio_com_ceis	Matserv incompatível	Baixa quantidade
0.0	False	False	False	False
0.0	False	False	False	False
1.0	False	False	False	False
0.0	False	False	False	False
0.0	False	False	False	False
IndSocioLaranja	IndQtdDiasInferior30	Mais de um fornecedor	NaoLoja	Loja
True	False	True	0	1
False	False	False	1	0
False	False	False	1	0
False	False	False	1	0
False	False	False	0	0

O modelo base possui dez colunas, sendo oito correspondentes a tipologias levantadas pelo ODP:

- “Mais de 10 subclasses” corresponde à tipologia 5 (“Fornecedores com grande quantidade de subclasses de CNAE”);
- “EPP/ME acima do limite” corresponde à tipologia 3 (“EPP/ME com valores totais homologados acima do limite”);
- “ind_socio_com_ceis” corresponde à tipologia 9 (“Sócios com empresas no CEIS”);
- “Matserv Incompatível” corresponde à tipologia 10 (“Compatibilidade entre atividade econômica e produto/serviço fornecido”);

- “Baixa quantidade”: corresponde à tipologia 1 (“Baixa quantidade de funcionários”);
- “IndSocioLaranja” corresponde à tipologia 2 (“Sócios com baixa qualificação ou remuneração”);
- “IndQtdDiasInferior30” corresponde à tipologia 6 (“Quantidade de dias entre abertura e data da compra inferior a 30”);
- “Mais de um fornecedor” corresponde à tipologia 4 (“Mais de um fornecedor no mesmo endereço”).

Já as duas últimas colunas são derivadas da coluna “classificacao”, que possui dados categóricos derivados da saída da rede convolucional: valores “Loja”, “NaoLoja” e “Indefinido”, sendo este último valor excluído, pois deriva dos outros dois valores e poderia trazer problemas de multicolinearidade no modelo. Desta forma, as colunas “Loja” e “NaoLoja” foram mantidas no modelo por não serem perfeitamente colineares (devido à existência da variável “Indefinido”).

O resultado do ajuste do modelo pode ser visto na Tabela 4.6. Os coeficientes relativos às saídas do modelo convolucional são ambos negativos (-0.0231 para “NaoLoja” e -0.1738 para “Loja”), porém somente o coeficiente de “Loja” possui significância estatística (p-valor < 0.05), muito provavelmente por termos uma grande correlação entre as últimas duas colunas, “Loja” e “NaoLoja”. De todo modo, o modelo nos diz que uma classificação de um cenário como comercial tende a diminuir o risco de um fornecedor ser punido, o que é coerente com os testes estatísticos realizados na seção anterior.

Quanto aos coeficientes das demais variáveis da regressão, podemos verificar que baixa quantidade de funcionários (“Baixa quantidade”, com coeficiente -0.3258) e material ou serviços incompatíveis com a atividade econômica (“Matserv Incompatível”, com coeficiente -1.2854) diminuem o risco de punição do fornecedor. A baixa quantidade de funcionários pode estar relacionada a valores de contratos mais baixos e, portanto, de menor risco para a Administração, o que justificaria o coeficiente negativo. Entretanto, a presença de um coeficiente negativo para “Matserv incompatível” não é intuitiva, pois se esperava que este atributo aumentasse o risco do fornecedor. Há que se considerar que a construção desta tipologia utilizou como base uma tabela de correspondências entre materiais e serviços do sistema ComprasNet SIASG e Classificação de Atividades Econômicas compatíveis que foi construída manualmente, sujeita portanto a subjetividades de interpretação e a erros de preenchimento. Estudos mais aprofundados do relacionamento desta tipologia com a variável dependente devem ser realizados em trabalhos futuros.

De acordo com o modelo, as variáveis “IndSocioLaranja”, “EPP/ME acima do limite”, “Mais de 10 subclasses”, “IndQtdDiasInferior30” e “ind_socio_com_ceis” possuem co-

eficientes positivos e portanto estão relacionadas a um aumento de risco de punição do fornecedor, o que era esperado. A variável relacionada a mais de um fornecedor em um mesmo endereço (“Mais de um fornecedor”) não possui coeficiente com significância estatística (p -valor ≥ 0.5) para que seja possível analisar seu relacionamento com o aumento ou diminuição do risco do fornecedor.

Tabela 4.6: Variáveis do modelo base.

Variável	Coefficiente	P-valor do coeficiente
Constante do modelo	-0.1164	0.073
Baixa quantidade	-0.3258	0.000
IndSocioLaranja	0.1575	0.000
EPP/ME acima do limite	1.0811	0.000
Mais de um fornecedor	0.0800	0.153
Mais de 10 subclasses	0.7890	0.000
IndQtdDiasInferior30	1.0329	0.000
ind_socio_com_ceis	1.2234	0.000
Matserv Incompativel	-1.2854	0.000
NaoLoja	-0.0231	0.730
Loja	-0.1738	0.008

Por utilizar apenas variáveis binárias, esta regressão torna-se pobre em informação e conseqüentemente em valor preditivo no conjunto de teste, conforme demonstrado pelas métricas de F-measure de 0.68 e Área sobre a curva ROC de 0.66, conforme Tabela 4.7 e Figura 4.2. Na Tabela 4.7, o Suporte é a quantidade de elementos de Fornecedores Punidos, Não Punidos e Totais considerados para cálculo das métricas de desempenho. Na Curva ROC da Figura 4.2, a circunferência laranja indica a taxa de verdadeiros positivos e falsos positivos atual no limite de 50% (ou seja, o limite onde valores na saída da regressão superiores a 50% indicam fornecedores da classe Punidos).

Tabela 4.7: Métricas de precisão, abrangência (*recall*) e F-Measure do modelo base.

	Precisão	Recall	F-Measure	Suporte
Fornecedor não punido	0.78	0.76	0.77	7033
Fornecedor punido	0.46	0.48	0.47	2930
Média / Total	0.69	0.68	0.68	9963

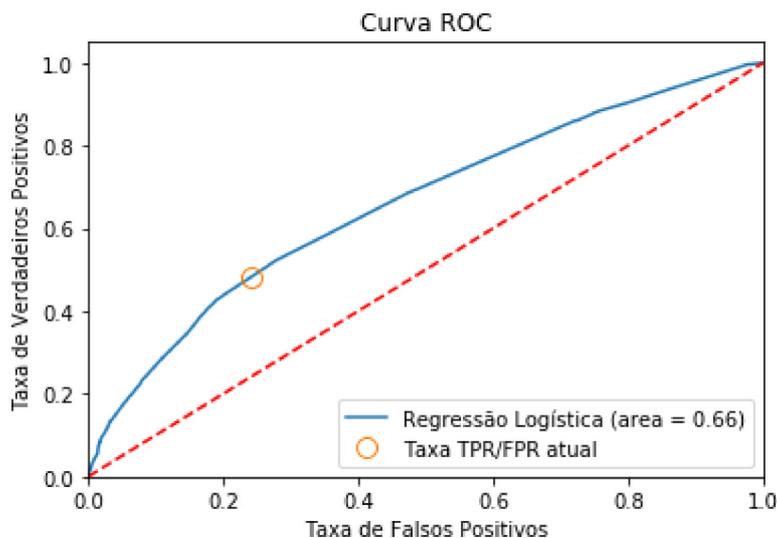


Figura 4.2: Curva ROC do modelo base.

4.3.2 Modelo Quantitativo

Se utilizarmos informações mais quantitativas relacionadas às tipologias podemos ajustar um modelo com melhor valor preditivo. Desta forma, optamos por utilizar as variáveis mostradas na Tabela 4.8.

Este modelo possui 12 variáveis relacionadas às seguintes tipologias:

- “qtd_media” é a quantidade média de funcionários da tipologia 1 (“Baixa quantidade de funcionários”);
- “Socio laranja” é a quantidade de sócios com baixa qualificação ou remuneração da tipologia 2 (“Sócios com baixa qualificação ou remuneração”);
- “Valor Total Homologado em 2017”, “Micro Empresa”, “Não Informado”, “Pequena Empresa” são variáveis de valor e porte da empresa derivadas da tipologia 3 (“EPP/ME com valores totais homologados acima do limite em 2017”);
- “qtd_subclasses_cnae” é a quantidade de subclasses de atividade econômica da tipologia 5 (“Fornecedores com grande quantidade de subclasses de CNAE”);
- “Qtd dias inferior a 30” é a quantidade exata de dias entre a abertura da empresa e a data da compra da tipologia 6 (“Quantidade de dias entre abertura e data da compra inferior a 30”);
- “qtd_fornecedores” é a quantidade de fornecedores no mesmo endereço da tipologia 4 (“Mais de um fornecedor no mesmo endereço”);

Tabela 4.8: Campos derivados das tipologias utilizadas para ajuste do modelo quantitativo.

qtd_media	Socio Laranja	Valor Total Homologado em 2017	Micro Empresa
96.0	0.0	100000.0	0
34.0	0.0	100000.0	0
28.0	0.0	100000.0	0
23.0	0.0	100000.0	0
25.0	0.0	76100.0	0
Não Informado	Pequena Empresa	qtd_subclasses_cnae	Qtd dias inferior a 30
0	0	5.0	0.0
0	0	5.0	0.0
0	0	5.0	0.0
0	0	5.0	0.0
1	0	5.0	0.0
qtd_fornecedores	socio_com_ceis	VALIDO	prob_loja
1.0	0.0	2.0	0.993297
1.0	0.0	7.0	0.999977
1.0	0.0	1.0	0.970694
1.0	0.0	1.0	0.926276
1.0	0.0	3.0	0.999390

- “socio_com_ceis” é a quantidade de sócios que possuem empresas no CEIS da tipologia 9 (“Sócios com empresas no CEIS”);
- “VALIDO” é a quantidade de produtos ou serviços compatíveis com a atividade econômica do fornecedor da tipologia 10 (“Compatibilidade entre atividade econômica e produto/serviço fornecido”);
- “prob_loja” é a probabilidade do cenário do fornecedor ser comercial, informação extraída da rede convolucional.

Devido à grande diferença entre as grandezas envolvidas nas variáveis deste modelo (bastando para isso observar a diferença entre os valores das colunas “Valor Total Homologado em 2017” e “qtd_media” em comparação com o restante das variáveis na Tabela 4.8), optamos por normalizar todas as colunas antes do ajuste deste modelo. O método de normalização escolhido foi a normalização 0-1, que é feita subtraindo das variáveis seu valor mínimo e dividindo o resultado pela sua amplitude (valor máximo - valor mínimo), de modo que todas as variáveis se encontrem dentro do intervalo [0,1].

O resultado do ajuste desse modelo pode ser visto na Tabela 4.9. Mais uma vez, vemos que o coeficiente relacionado à probabilidade do fornecedor ser comercial é negativo, o que indica que a classificação de um fornecedor neste tipo de cenário está ligada à diminuição do risco do fornecedor ser punido.

Quanto aos outros coeficientes, podemos verificar que as variáveis “Não Informado” (coeficiente -2.693) e “qtd_subclasses_cnae” (coeficiente -1.7381), por possuírem coefici-

entes negativos, também diminuem o risco de punição do fornecedor. São resultados de difícil interpretação, podendo representar correlações espúrias que poderiam ser melhor explicadas ao adicionarmos mais variáveis ao modelo. Interessante verificar que a variável relacionada à tipologia 5 (“Quantidade de subclasses de atividade econômica superior a 10”) possui coeficiente positivo (aumenta o risco do fornecedor) no modelo base com variáveis categóricas, porém tem coeficiente negativo (diminui o risco do fornecedor) no modelo que analisa variáveis quantitativas. Isto pode indicar que tanto a presença de baixa quantidade de atividades econômicas como o excesso delas (superior a 10) podem ser indicadores de aumento de risco do fornecedor.

A maioria das variáveis da regressão na Tabela 4.9, entretanto, indicam aumento de risco de punição do fornecedor por possuírem coeficientes positivos. O aumento de risco de punição do fornecedor com o aumento das variáveis “qtd_media”, “Socio Laranja”, “Valor total Homologado em 2017”, “qtd_fornecedores”, “Qtd dias inferior a 30” e “socio_com_ceis” é algo esperado. Já o aumento de risco do fornecedor com o aumento das variáveis “Micro Empresa” e “Pequena Empresa” é de interpretação menos óbvia, assim como o aumento do risco do fornecedor pelo aumento de itens de material ou serviço considerados compatíveis com a atividade econômica exercida (variável “VALIDO”), pois se esperava a diminuição do risco neste último caso. A interpretação dessas três últimas variáveis, em especial a variável “VALIDO”, construída a partir de uma correspondência manual entre materiais e serviços e atividades econômicas (sujeita a subjetividades e erros), merece ser aprofundada em estudos posteriores.

Tabela 4.9: Variáveis do modelo quantitativo.

Variável	Coefficiente	P-valor do coeficiente
Constante do modelo	-0.7689	0.000
qtd_media	15.8415	0.000
Socio Laranja	4.1915	0.000
Valor Total Homologado em 2017	17.8808	0.000
Micro Empresa	0.3900	0.000
Não Informado	-2.6930	0.000
Pequena Empresa	0.5276	0.000
qtd_fornecedores	1.2099	0.114
qtd_subclasses_cnae	-1.7381	0.000
Qtd dias inferior a 30	15.1842	0.000
socio_com_ceis	11.1347	0.000
VALIDO	21.1931	0.000
prob_loja	-0.1512	0.002

Se compararmos este modelo com modelos similares que incluem qualquer outra das variáveis relacionadas à rede convolucional (como “Loja”, “NaoLoja” ou “Indefinido”, ver Tabela 4.10), vemos que os coeficientes podem se inverter, mas se mantêm logicamente consistentes. No caso da variável “Loja” ele se mantém negativo, e no caso “NaoLoja” o coeficiente possui naturalmente o sinal positivo. No caso da variável “Indefinido”, o p-valor alto encontrado não permite fazer afirmações sobre seu coeficiente.

Tabela 4.10: Comparando o modelo com outras variáveis da rede convolucional.

Tipo de Modelo	Coefficiente da variável da CNN	P-valor da variável da CNN
Modelo quantitativo original (com prob_loja)	-0.1512	0.002
Modelo + Loja	-0.1406	0.000
Modelo + NaoLoja	0.1289	0.000
Modelo + Indefinido	0.1141	0.105

Sem necessidade de parâmetros de regularização, vemos que as métricas de desempenho do modelo mostram melhora significativa nas estimativas de F-Measure (0.76) e Área sobre a curva ROC (0.8), conforme Tabela 4.11 e Figura 4.3, fruto não propriamente do aumento do número de variáveis do modelo, mas sim da quantidade de informação disponível em cada variável. Ao utilizarmos método de seleção de variáveis³, não houve redução de nenhuma das 12 variáveis utilizadas.

Tabela 4.11: Métricas de precisão, abrangência (recall) e F-Measure do modelo quantitativo.

	Precisão	Recall	F-Measure	Suporte
Fornecedor não punido	0.84	0.81	0.83	7033
Fornecedor punido	0.58	0.63	0.61	2930
Média / Total	0.76	0.76	0.76	9963

É importante verificar que, embora haja significância estatística para o coeficiente da variável de saída do modelo convolucional (“prob_loja”), a sua retirada afeta muito pouco o desempenho do modelo, conforme Tabela 4.12 e Figura 4.4.

³Método de seleção de variáveis utilizando o algoritmo *Stepwise Forward Selection* com a métrica de Critério de Informação de Akaike (AIC).

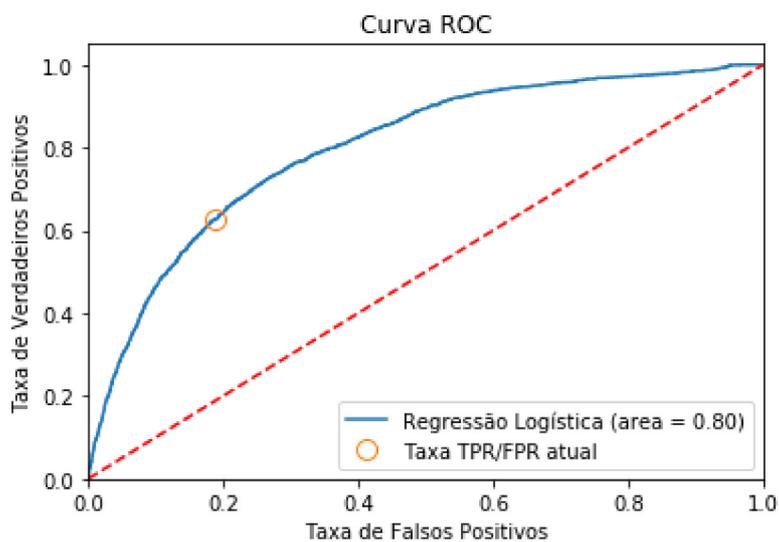


Figura 4.3: Curva ROC do modelo quantitativo.

Tabela 4.12: Métricas de precisão, abrangência (recall) e F-Measure do modelo quantitativo sem a variável “prob_loja”.

	Precisão	Recall	F-Measure	Suporte
Fornecedor não punido	0.84	0.82	0.83	7033
Fornecedor punido	0.58	0.62	0.60	2930
Média / Total	0.76	0.76	0.76	9963

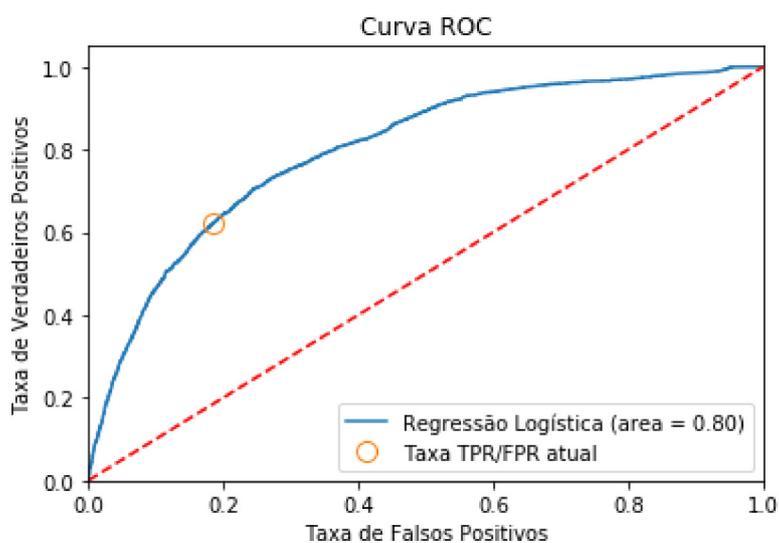


Figura 4.4: Curva ROC do modelo quantitativo sem a variável “prob_loja”.

4.3.3 Modelos de Regressão por CNAE

Nesta última análise, verificamos se os efeitos da saída do modelo convolucional possuem influência maior para alguma classe específica de atividade econômica do fornecedor.

Para isso, primeiramente fizemos uma análise simples, comparando modelos de regressão logística para cada uma das seções da Classificação Nacional de Atividades Econômicas (CNAE) utilizando a variável de punição suspensiva do fornecedor como variável dependente e somente a variável “prob_loja” (saída da rede convolucional) como variável dependente.

Os coeficientes e p-valores da variável “prob_loja” para cada um desses modelos, com exceção da Seção U (“Organismos internacionais e outra instituições extraterritoriais”), que não possui registro de punição de fornecedores, estão na Tabela 4.13.

Podemos ver que sete seções da CNAE (E, G, O, P, Q, R e S) apresentaram valores estatisticamente significativos para o coeficiente da variável “prob_loja”.

Entretanto, este relacionamento pode ter ocorrido devido a outras variáveis do modelo, que ignoramos nesta análise preliminar. Devemos portanto avançar na análise e verificar se estes valores se mantêm significativos ao analisarmos modelos de regressão logística destas sete seções considerando todas as 12 variáveis independentes do modelo quantitativo da seção anterior e não somente a variável “prob_loja”. Os resultados desses modelos (mostrando apenas os coeficientes e p-valores de “prob_loja”) estão na Tabela 4.14.

Conforme Tabela 4.14, as seções de atividade econômica que mantiveram significância estatística mesmo quando consideradas as outras variáveis do modelo foram as Seções “O” (Administração Pública, Defesa e Seguridade Social), “Q” (Saúde Humana e Serviços Sociais) e “S” (Outras Atividades de Serviços).

A Seção “O” (Administração Pública, Defesa e Seguridade Social) oferece em maior parte serviços de recolhimento de taxas e de assinatura e publicação de periódicos oficiais. Já a Seção “Q” (Saúde Humana e Serviços Sociais) oferece em maior parte serviços de assistência médica/hospitalar, análises clínicas e exames médicos. A Seção “S” (Outras Atividades de Serviços) agrega tipos de serviços não classificados nas outras seções de atividade econômica.

Nessas seções, o peso do coeficiente de saída da rede convolucional é significativamente maior que no modelo completo (-2.7510, -0.4687 e -0.6431 para as Seções “O”, “Q” e “S”, respectivamente, contra apenas -0.1512 do modelo de regressão com os dados completos), o que significa que, para estas classes de atividade econômica, a presença de um fornecedor em cenário comercial está relacionada mais fortemente à diminuição de risco de punição.

Tabela 4.13: Modelos de regressão com “prob_loja” por seção da CNAE.

Seção de Atividade Econômica (CNAE)	Coefficiente de prob_loja	P-valor de prob_loja
A - Agricultura, Pecuária, Produção Florestal, Pesca E Aquicultura	0.6881	0.161
B - Indústrias Extrativas	0.9051	0.443
C - Indústrias de Transformação	-0.0667	0.424
D - Eletricidade e Gás	0.6020	0.271
E - Água, Esgoto, Atividades de Gestão de Resíduos e Descontaminação	-0.6101	0.005
F - Construção	0.1086	0.206
G - Comércio; Reparação de Veículos Automotores e Motocicletas	-0.2436	0.000
H - Transporte, Armazenagem e Correio	0.0914	0.560
I - Alojamento e Alimentação	0.2111	0.306
J - Informação e Comunicação	0.0109	0.933
K - Atividades Financeiras, de Seguros e Serviços Relacionados	0.3384	0.344
L - Atividades Imobiliárias	-0.2434	0.298
M - Atividades Profissionais, Científicas e Técnicas	0.0802	0.452
N - Atividades Administrativas e Serviços Complementares	-0.1305	0.162
O - Administração Pública, Defesa e Seguridade Social	-2.0433	0.000
P - Educação	-0.3286	0.038
Q - Saúde Humana e Serviços Sociais	-0.2723	0.000
R - Artes, Cultura Esporte e Recreação	0.8291	0.002
S - Outras Atividades de Serviços	-0.2856	0.033
U - Organismos Internacionais e Outras Instituições Extraterritoriais	-	-

Tabela 4.14: Modelos de regressão com 12 variáveis do modelo quantitativo por seção da CNAE.

Seção de Atividade Econômica (CNAE)	Coefficiente de prob_loja	P-valor de prob_loja
E - Água, Esgoto, Atividades de Gestão de Resíduos e Descontaminação	-0.5128	0.079
G - Comércio; Reparação de Veículos Automotores e Motocicletas	-0.1274	0.051
O - Administração Pública, Defesa e Seguridade Social	-2.7510	0.002
P - Educação	-0.2653	0.204
Q - Saúde Humana e Serviços Sociais	-0.4687	0.000
R - Artes, Cultura, Esporte e Recreação	0.2002	0.578
S - Outras Atividades de Serviços	-0.6431	0.000

Capítulo 5

Conclusões e Trabalhos Futuros

O trabalho de ajuste do modelo Wide Resnet18 a partir de imagens de fornecedores nos permitiu concluir que é possível ajustar com sucesso modelos de redes convolucionais para diferenciar cenários entre comerciais e não comerciais.

Além disso, nosso trabalho em Wacker, Ferreira e Ladeira [38] mostra que a escolha por uma orientação específica nas imagens baixadas para treinamento (seja ela “Norte”, “Sul”, “Leste” ou “Oeste”) não afeta o desempenho dos modelos treinados, podendo portanto ser utilizada qualquer uma destas orientações, ou todas em conjunto, para realizar o ajuste fino da rede convolucional. Desta forma, utilizamos todas as imagens anotadas entre cenários comerciais e não comerciais do nosso conjunto de treino, independentemente da orientação baixada, para realizar o ajuste fino da rede Wide Resnet18, visto que esta abordagem maximiza a quantidade de imagens utilizadas para alimentar o modelo.

Outra conclusão interessante diz respeito aos atributos extraídos da penúltima camada da rede (Conv17), que possuem a capacidade de reconhecer objetos como carros e complexos residenciais. Trabalhos futuros podem utilizar apenas as saídas que interessam para a classificação final, sem necessidade de utilizar todos os 512 atributos de saída da penúltima camada, podendo levar a ganhos de eficiência na classificação dos cenários.

A análise de modelos simples de regressão logística mostra que de fato podem existir relações entre as classes descobertas por redes convolucionais e as tipologias de risco levantadas pelos auditores da CGU. Apesar de simples, o modelo de regressão logística completo envolvendo 12 variáveis quantitativas possui um bom desempenho (F-Measure de 0.76 e área sobre a curva ROC de 0.80). Este desempenho é inferior ao modelo proposto por Sales para a CGU [3] (F-Measure de 0.86), porém é importante observar que nosso modelo completo utilizou apenas 12 variáveis contra 29 do modelo de Sales.

No caso específico, a classificação de um cenário comercial está relacionada à ligeira diminuição de risco de punição suspensiva de um fornecedor (o coeficiente de `prob_loja` na regressão logística foi de -0.15), e este relacionamento possui significância estatística

em todos os testes que fizemos (p -valor < 0.05 nos testes Qui-Quadrado e através dos modelos de Regressão Logística).

Não se pode concluir, entretanto, que essa relação possui algum valor preditivo, já que a exclusão deste tipo de variável não piora significativamente o desempenho do modelo de regressão logística testado (F-Measure se manteve em torno de 0.76 e a área sobre a curva ROC se manteve em torno de 0.80 mesmo após a exclusão da variável `prob_loja`).

De fato, a interpretação que tivemos de cenários comerciais como simples lojas físicas de rua não se mostrou adequada ao nosso contexto. Em trabalhos futuros, devemos considerar outros tipos de cenários comerciais válidos para anotação, tais como shoppings e prédios comerciais, que podem ser mais representativos de fornecedores típicos da Administração Pública Federal e, portanto, podem vir a apresentar coeficientes com maior valor preditivo para modelos de riscos em geral.

Para fornecedores de algumas seções de atividade econômica, contudo, a classificação de um cenário comercial está mais fortemente ligada à diminuição do risco de punição do fornecedor. Entre elas, podemos citar as Seções “O” (Administração Pública, Defesa e Seguridade Social), “Q” (Saúde Humana e Serviços Sociais) e “S” (Outras Atividades de Serviços). Para estas seções, temos coeficientes de regressão logística iguais a -2.75, -0.46 e -0.64, respectivamente, todos valores muito superiores em módulo ao coeficiente de `prob_loja` (-0.15).

Dentre as limitações existentes neste trabalho estão a falta de disponibilidade de imagens para todos os endereços de fornecedores, a desatualização de conjuntos de imagens baixadas e a imprecisão presente ao buscarmos localizações através de mera descrição textual de endereços (e não de coordenadas exatas) de fornecedores.

Trabalhos futuros também devem levar em conta que a própria ausência de imagem de um fornecedor no Street View pode ser um indicativo de que este está localizado em local atípico de pouca circulação. Estudos semelhantes ao deste trabalho podem então ser feitos a fim de avaliar se a ausência dessas imagens é um atributo relevante em um modelo de risco mais geral. Quanto à desatualização das imagens e imprecisão de endereços textuais, acreditamos que estes problemas possam ser resolvidos se for possível exigir que os licitantes informem as coordenadas de suas sedes, bem como fotos atualizadas do local, se necessário.

O que se deve buscar em trabalhos futuros é o ajuste da rede convolucional para reconhecimento de novos tipos de cenários de fornecedores, inclusive considerando os diversos tipos de classes de atividade econômica existentes, pois um conjunto maior e mais diverso de informações pode trazer novos relacionamentos que possam de fato melhorar as capacidades preditivas dos modelos de risco atualmente em uso na CGU.

Uma forma de se encontrar novos tipos de cenários é realizando a clusterização (ou

agrupamento) das imagens obtidas, com o objetivo de encontrar grupos de imagens homogêneas que possam servir de base para anotação de novos tipos de cenários.

O exemplo da Figura 5.1 mostra uma clusterização experimental feita em um subconjunto das imagens obtidas neste trabalho. Neste caso, foi utilizado o algoritmo de clusterização aglomerativa com a métrica da similaridade dos cossenos¹. Para encontrar elementos com uma boa separação no conjunto de dados, normalmente utilizamos a medida da Silhueta [41], calculada em nosso contexto para cada imagem i por:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Onde $a(i)$ é a distância média da imagem i para as imagens de seu próprio grupo e $b(i)$ é a distância média da imagem i para as imagens do grupo mais próximo.

A métrica da Silhueta verifica o quanto um determinado elemento está de fato próximo do cluster (ou grupo) a que pertence (baixo $a(i)$) enquanto se encontra suficientemente distante dos outros clusters (alto $b(i)$). A métrica varia entre -1 e 1. Fica próxima de 1 se de fato a clusterização é uma boa separação para o elemento (baixo $a(i)$ e alto $b(i)$) e próxima de -1 caso seja uma separação ruim (alto $a(i)$ e baixo $b(i)$).

Desta forma, uma medida da qualidade da separação de um cluster pode ser dada pela Silhueta Média desse cluster, ou seja, para um cluster C , sua silhueta média é:

$$S_c = \sum_{i \in C} \frac{s(i)}{|C|}$$

Embora a métrica da Silhueta seja comumente utilizada para encontrar clusters com uma boa separação no conjunto de dados, no exemplo dado ela não serviu bem para diferenciar grupos homogêneos (rodovias) de grupos mais heterogêneos, como podemos verificar na Figura 5.1. Entretanto, é possível observar na figura que o valor máximo da distância média entre as imagens de um mesmo cluster (denotado por $\max(a)$) pode ser uma melhor métrica para encontrar exemplos de cenários homogêneos em trabalhos futuros.

Por fim, ressaltamos a importância deste trabalho no âmbito da CGU, capaz de fomentar trabalhos relacionados à visão computacional e à utilização de imagens como fontes de dados. Um exemplo de trabalho relacionado é o de acompanhamento de obras de creches construídas pelo Programa ProInfância com recursos do Fundo Nacional de De-

¹De acordo com a documentação da biblioteca Scikit-learn [40], a clusterização aglomerativa é um método de clusterização que utiliza a abordagem *bottom-up*, ou seja, inicia com cada elemento em um cluster próprio e agrupa esses clusters em grupos maiores através de uma métrica de distância ou de similaridade. No exemplo, utilizamos a métrica de similaridade dos cossenos, que é o produto interno normalizado de dois vetores, calculado por $\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$.

Silhueta: 0,43 - Max(a): 0,26



Silhueta: 0,40 - **Max(a): 0,51**



Figura 5.1: Comparação de dois clusters de imagens de fornecedores baixadas do Google Street View utilizando as métricas de silhueta e distância máxima intra-cluster (Max(a)).

Fonte: Google Stree View.

envolvimento da Educação (FNDE) através de imagens baixadas do Google Earth (já em andamento, ver Figura 5.2).

Além deste trabalho, já em andamento, podemos citar também trabalhos futuros que envolvem as mesmas técnicas de ajuste fino de redes convolucionais mostradas neste projeto.

Um exemplo de trabalho desta natureza, previsto para o início de 2019, é o de detecção de rachaduras em estradas federais com Google Street View e uma rede convolucional pré-treinada, extraída de Zhang et al. [42]. Na Figura 5.3, é possível identificar uma rachadura visível em trecho de pista que posteriormente cedeu no centro de Brasília/DF. Um projeto para ajuste de redes convolucionais para identificar tais rachaduras em estradas terá início na CGU em 2019.



Figura 5.2: Exemplo de creche do programa ProInfância sendo construída ao longo dos anos.

Fonte: Google Earth.



Figura 5.3: Exemplo de rachadura visível em trecho de pista no centro de Brasília/DF.

Fonte: Google Street View.

Referências

- [1] N. Dassen and J. C. Vieyra, “Open Government and Targeted Transparency,” *Trends and Challenges for Latin America and the Caribbean, Washington, DC, Banco Interamericano de Desarrollo*, 2012. 1
- [2] E. v. d. D. d. Willebois, J. C. Sharman, R. Harrison, J. W. Park, and E. Halter, *The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It*. World Bank Publications, Nov. 2011. Google-Books-ID: WdTJ6LPhBxYC. 2, 6
- [3] L. J. Sales, “Proposta de modelo de classificação do risco de contratos públicos,” Master’s thesis, Universidade de Brasília, Sep. 2016. 3, 18, 51
- [4] U. Qidwai and C. H. Chen, *Digital Image Processing: An Algorithmic Approach with MATLAB*. CRC Press, Apr. 2009. 7
- [5] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, p. 436, 2015. 8, 10, 11
- [6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. 8, 9
- [7] J. Grus, *Data Science from Scratch: First Principles with Python*. O’Reilly Media, Inc., 1st ed., 2015. 10
- [8] Y. Ma, Z. Xiang, Q. Du, and W. Fan, “Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning,” *International Journal of Hospitality Management*, vol. 71, pp. 120–131, 04 2018. 11
- [9] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015. 12
- [10] Y. Bengio, “Practical recommendations for gradient-based training of deep architectures,” in *Neural networks: Tricks of the trade*, pp. 437–478, Springer, 2012. 12
- [11] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” in *Advances in neural information processing systems*, pp. 3320–3328, 2014. 12
- [12] T. Fawcett, “An Introduction to ROC analysis,” *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006. 15, 16, 17

- [13] T. Fawcett and F. Provost, “Adaptive fraud detection,” *Data mining and knowledge discovery*, vol. 1, no. 3, pp. 291–316, 1997. 17
- [14] J. Neville and D. Jensen, “Iterative classification in relational data,” in *Proc. AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, pp. 13–20, 2000. 17
- [15] R. J. Bolton and D. J. Hand, “Statistical fraud detection: A review,” *Statistical science*, pp. 235–249, 2002. 18
- [16] M. Albashrawi, “Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015,” *Journal of Data Science*, vol. 14, no. 3, pp. 553–569, 2016. 18
- [17] C. T. Spathis, “Detecting false financial statements using published data: some evidence from greece,” *Managerial Auditing Journal*, vol. 17, no. 4, pp. 179–191, 2002. 18
- [18] T. Sun and L. J. Sales, “Predicting Public Procurement Irregularity: An Application of Neural Networks,” *Journal of Emerging Technologies in Accounting: Spring 2018*, vol. 15, no. 1, pp. 141–154, 2018. 19
- [19] A. K. Griffith, “Mathematical models for automatic line detection,” *Journal of the ACM (JACM)*, vol. 20, no. 1, pp. 62–80, 1973. 19
- [20] A. K. Griffith, “Edge detection in simple scenes using a priori information,” *IEEE Transactions on Computers*, vol. 100, no. 4, pp. 371–381, 1973. 19
- [21] A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001. 19
- [22] J. Zou, W. Li, C. Chen, and Q. Du, “Scene classification using local and global features with collaborative representation fusion,” *Information Sciences*, vol. 348, pp. 209–226, 2016. 19
- [23] L. Fei-Fei and P. Perona, “A bayesian hierarchical model for learning natural scene categories,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 524–531, IEEE, 2005. 19
- [24] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, “Evaluating bag-of-visual-words representations in scene classification,” in *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pp. 197–206, ACM, 2007. 19
- [25] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, vol. 2, pp. 2169–2178, IEEE, 2006. 20, 24

- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012. 20
- [27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *arXiv:1409.0575 [cs]*, Sept. 2014. arXiv: 1409.0575. 20
- [28] T. Harada and Y. Kuniyoshi, “Graphical gaussian vector for image categorization,” in *Advances in Neural Information Processing Systems*, pp. 1547–1555, 2012. 20
- [29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv:1512.03385 [cs]*, Dec. 2015. arXiv: 1512.03385. 20, 21, 24
- [30] S. Zagoruyko and N. Komodakis, “Wide residual networks,” *arXiv preprint arXiv:1605.07146*, 2016. 22
- [31] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva, “Places: An image database for deep scene understanding,” *arXiv preprint arXiv:1610.02055*, 2016. 22, 23, 24, 28, 30
- [32] M. A. R. Alif, S. Ahmed, and M. A. Hasan, “Isolated bangla handwritten character recognition with convolutional neural network,” in *Computer and Information Technology (ICCIT), 2017 20th International Conference of*, pp. 1–6, IEEE, 2017. 23
- [33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015. 23
- [34] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, “Sun database: Large-scale scene recognition from abbey to zoo,” in *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pp. 3485–3492, IEEE, 2010. 24
- [35] A. Quattoni and A. Torralba, “Recognizing indoor scenes,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 413–420, IEEE, 2009. 24
- [36] G. Patterson and J. Hays, “Sun attribute database: Discovering, annotating, and recognizing scene attributes,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2751–2758, IEEE, 2012. 24
- [37] R. Wirth and J. Hipp, “Crisp-dm: Towards a standard process model for data mining,” in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pp. 29–39, 2000. 27
- [38] J. Wacker, R. P. Ferreira, and M. Ladeira, “Detecting fake suppliers using deep image features,” in *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 224–229, IEEE, 2018. 29, 30, 31, 35, 36, 37, 51

- [39] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, “Learning Deep Features for Scene Recognition using Places Database,” in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 487–495, Curran Associates, Inc., 2014. 30
- [40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. 53
- [41] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987. 53
- [42] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, “Road crack detection using deep convolutional neural network,” in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3708–3712, IEEE. 54