



**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 qualificação do  
Mestrado Profissional em Computação Aplicada

Orientador  
Prof. Dr. Marcelo Ladeira

Brasília  
2017



# 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 de origem ilícita. Este trabalho propõe a aplicação de um classificador de imagens para detecção de tais fornecedores, utilizando redes neurais convolucionais (CNN em inglês), mais especificamente a arquitetura AlexNet. As imagens serão obtidas através de pesquisas dos endereços dos fornecedores pela API do Google Street View, e a classificação resultante deverá complementar outros atributos já obtidos para detecção de fornecedores de fachada, de forma a melhorar a efetividade na sua detecção por parte dos auditores do Ministério da Transparência e Controladoria-Geral da União (CGU). Até o presente momento, foram classificados cerca de 15 mil cenários de fornecedores em contratos ativos com o Governo Federal até setembro de 2016, obtidos a partir de cruzamento de dados entre as bases de Compras Governamentais (Comprasnet SIASG) e da Receita Federal (cenários mais recentes devem ser classificados ao longo do trabalho). Os resultados preliminares obtidos por classificação através de uma rede AlexNet pré-treinada mostram a necessidade de criação de novas classes de cenários adaptadas ao contexto nacional. Para obtermos estas classes, propomos a aplicação de algoritmos de clusterização nos atributos aprendidos pela rede, extraídos de sua última camada. As novas classes obtidas serão utilizadas para realizar o ajuste fino da rede AlexNet para classificações futuras, através de seu retreinamento completo ou ajuste fino da rede.

**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 proposes to apply image classification to detect such organizations, through the use of Convolutional Neural Networks (CNN), namely the AlexNet architecture. The images are obtained by address search in Google Street View API, and the resulting classification will be further used along with other features to detect front companies in order to help the auditors from the Ministry of Transparency and Office of the Comptroller General (CGU, in Portuguese). To this moment, we applied classification to almost 15 thousand suppliers scenes with active contracts with the Brazilian Government until September 2016, obtained through data matching between the Government Purchases database and the Brazilian Federal Revenue Office database (more recent scenes should be added as this work progresses). Preliminary results with a pre-trained AlexNet CNN show the need for developing new scene classes more suited to the Brazilian context. In order to do this, we propose to apply clustering algorithms in features extracted from the last fully-connected layer of this net. The classes obtained will be used to fine-tune the AlexNet CNN for future classification, through the use of training from scratch or fine tuning techniques.

**Keywords:** Front companies, Scene Recognition, Convolutional Neural Networks

# Sumário

<b>1 Definição do Problema</b>	<b>1</b>
1.1 A CGU e o Observatório da Despesa Pública . . . . .	1
1.2 Fornecedores de Fachada . . . . .	2
1.3 Justificativa do Tema . . . . .	2
1.4 Hipóteses de Pesquisa . . . . .	3
1.5 Objetivos . . . . .	4
1.5.1 Objetivo Geral . . . . .	4
1.5.2 Objetivos Específicos . . . . .	4
1.6 Contribuição Esperada . . . . .	5
<b>2 Revisão do estado da arte</b>	<b>6</b>
2.1 Detecção de Fornecedores de Fachada . . . . .	6
2.1.1 Modelos utilizados pela CGU . . . . .	8
2.2 Reconhecimento de Cenários . . . . .	8
2.3 Bancos de Dados de Cenários . . . . .	11
<b>3 Solução Proposta</b>	<b>13</b>
3.1 Metodologia . . . . .	15
3.1.1 Realizar pesquisa bibliográfica . . . . .	15
3.1.2 Obter as imagens dos fornecedores . . . . .	15
3.1.3 Obter classes de cenários . . . . .	16
3.1.4 Realizar ajuste fino da CNN . . . . .	17
3.1.5 Realizar a validação e teste dos modelos . . . . .	18
3.1.6 Realizar implantação do modelo final . . . . .	18
3.2 Experimentos Realizados . . . . .	19
3.3 Cronograma . . . . .	20
<b>4 Resultados Esperados</b>	<b>22</b>
<b>Referências</b>	<b>23</b>

# Lista de Figuras

2.1	Evolução das dimensões da imagem através da AlexNet CNN. . . . .	11
3.1	As cinco categorias de cenários mais comuns encontradas pela rede AlexNet.	19
3.2	Cenários classificados pela rede AlexNet. . . . .	21

# Lista de Tabelas

3.1 Cronograma . . . . .	21
--------------------------	----

# Lista de Abreviaturas e Siglas

**BoVW** *Bag-of-Visual-Words.*

**CGU** Ministério da Transparência e Controladoria-Geral da União.

**CNAE** Classificação Nacional de Atividades Econômicas.

**CNEFE** *Cadastro Nacional de Endereços para Fins Estatísticos.*

**CNN** Redes Neurais Convolucionais.

**Comprasnet SIASG** Sistema de Administração de Serviços Gerais.

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

**IBGE** *Instituto Brasileiro de Geografia e Estatística.*

**ILSVRC** *ImageNet Large Scale Visual Recognition Challenge.*

**LDA** Análise de Discriminantes Lineares.

**ODP** Observatório da Despesa Pública.

**RAIS** Relação Anual de Informações Sociais.

**RFB** Receita Federal do Brasil.

**Sistema APC** Análise Preventiva de Contratos.

**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 atípicos em relação às suas atividades e que possuem contrato com órgãos do Poder Executivo Federal, através da obtenção e classificação das imagens dos cenários onde estão localizados.

### 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 <sup>1</sup>, 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 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

---

<sup>1</sup>Decreto 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>).

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.

## 1.2 Fornecedores de Fachada

Dentre os temas monitorados pelo ODP, as licitações públicas ocupam um papel relevante. Licitações totalizando mais de R\$ 80 bilhões foram registradas no ano passado no Sistema de Administração de Serviços Gerais (Comprasnet SIASG), o principal sistema de licitações públicas do país<sup>2</sup>. 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.

## 1.3 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 atualmente 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. No ano passado, com base em dados do Ministério Público, Polícia Federal e Receita Federal, o jornal Estado de São Paulo<sup>3</sup> identificou 34 empresas investigadas pela operação consideradas de fachada ou “noteiras”<sup>4</sup>, 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

---

<sup>2</sup>De acordo com as informações disponíveis no Comprasnet SIASG, em 04 de agosto de 2017. Itens com valores maiores ou iguais a R\$ 1 bilhão (possíveis erros de preenchimento) foram excluídos.

<sup>3</sup>‘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>).

<sup>4</sup>Empresa “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.

são levantadas periodicamente através do Sistema de Análise Preventiva de Contratos da CGU (Sistema APC), tais como:

- quantidade de sócios;
- quantidade de funcionários;
- quantidade de possíveis sócios laranjas<sup>5</sup> (através do nível de escolaridade dos sócios);
- diferença entre a data da abertura da empresa e a data da compra pela unidade administrativa.

A quantidade de sócios e de possíveis sócios laranjas são obtidas através da base de dados da Receita Federal do Brasil (RFB). A quantidade de funcionários é obtida através dos dados da Relação Anual de Informações Sociais (RAIS). Já a informação da diferença entre a data de abertura da empresa e a data da compra pela unidade administrativa é obtida através do cruzamento entre as bases da RFB e do Comprasnet SIASG.

Estas 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 pelos sistemas e modelos da CGU, contribuindo para um aperfeiçoamento na detecção de fornecedores de fachada.

## 1.4 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 como atípica em relação às atividades que desenvolve.**  
Pretende-se verificar que a utilização de redes neurais convolucionais (CNN) pode

---

<sup>5</sup>No contexto deste trabalho, consideramos sócio laranja a pessoa apontada como parte do quadro societário da empresa apenas para criar uma “aparência” de legalidade, tendo ou não a pessoa conhecimento deste vínculo.

ser útil para identificar padrões nas imagens que indiquem localizações atípicas para os fornecedores, tais como lugares isolados ou pouco urbanizados.

- **a identificação de empresas em cenários atípicos em relação às atividades da empresa é um indicador relevante para determinar se ela é uma empresa de fachada.** Pretende-se verificar que a classificação de cenários pode gerar um atributo relevante para, juntamente com outros atributos já obtidos pela CGU (como, por exemplo, quantidade de empregados e nível de escolaridade dos sócios), seja possível melhorar a eficiência e a eficácia do trabalho dos auditores na identificação dessas empresas.

## 1.5 Objetivos

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

### 1.5.1 Objetivo Geral

Aperfeiçoar, no âmbito da CGU, os métodos para identificação de fornecedores de fachada e fornecedores de risco através da classificação e análise de imagens consideradas atípicas para fornecedores contratados pelo Governo Federal.

### 1.5.2 Objetivos Específicos

- ajustar modelos de CNN existentes para classificação de cenários (tais como AlexNet, GoogLeNet e ResNet) de forma a torná-los capazes de identificar de forma automatizada padrões atípicos em cenários de sedes de fornecedores do Governo Federal.
- pesquisar métodos eficientes de programação paralela e distribuída para a obtenção e classificação de grandes quantidades de imagens.
- verificar, por meio de técnicas de amostragem, se houve ganho de eficiência (medida através da diminuição de falso-positivos) na identificação de fornecedores de fachada através da inclusão da classificação automática de imagens quando comparada a abordagens anteriores que não utilizavam esta informação.
- caso haja ganho de eficiência na detecção de fornecedores de fachada através da inclusão da classificação automática de imagens, incluir esta funcionalidade no Sistema APC da CGU e no modelo de ranqueamento de contratos governamentais [3].

## 1.6 Contribuição Esperada

Como contribuição esperada após o fim do projeto, espera-se:

- aperfeiçoar os procedimentos para detecção de fornecedores de fachada através da inclusão de cenários classificados por CNN;
- integrar classes de cenários ao Sistema APC atualmente em uso na CGU;
- integrar classes de cenários ao modelo de ranqueamento de contratos governamentais da CGU [3];
- fomentar 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.

# Capítulo 2

## Revisão do estado da arte

Neste capítulo, apresentaremos uma revisão bibliográfica levando em conta a evolução do estado da arte em áreas consideradas importantes para este trabalho: Detecção de Fornecedores de Fachada e Reconhecimento de Cenários.

### 2.1 Detecção de 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 ou 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 listados, 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.

A detecção de fornecedores de fachada através de técnicas de mineração de dados, assim como técnicas de detecção de fraudes em geral, foi largamente influenciada pelos trabalhos de Fawcett e Provost em [4]. 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 [5] propuseram a utilização incremental de classificadores Bayesianos a partir de dados com alto grau de relacionamento, utilizando como exemplo a identificação 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 [6]: 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 [6], 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 [6], 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 [7] concluiu que as técnicas mais largamente utilizadas são Regressão Logística, Redes Neurais, Árvore de Decisão, Máquinas de Vetores de 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 [7] (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 [8], 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 fornecem declarações enganosas a respeito de sua real capacidade em cumprir suas obrigações no

fornecimento dos bens e serviços acordados.

Desta forma, de modo complementar à utilização de CNN para classificação de cenários destes fornecedores, propomos também a utilização das classes obtidas por estas redes como atributos em outros modelos, como Regressão Logística, Árvore de Decisão e Redes Neurais, que também utilizarão informações financeiras fornecidas por fornecedores do Governo Federal bem como dados financeiros obtidos ou inferidos das bases de dados da RFB, RAIS e ComprasNet SIASG. Os modelos de aprendizagem de máquina propostos foram escolhidos tendo como base os modelos mais utilizados na literatura para detecção de declarações financeiras falsas, de acordo com [7].

### **2.1.1 Modelos utilizados pela CGU**

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, 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 fornecedores, neste caso, é estimado via Regressão Logística, e os contratos são ranqueados para priorização das ações de auditoria.

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 para melhorar os resultados do Sistema APC da CGU, já citado no Capítulo 1.

## **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 [9][10].

O primeiro modelo amplamente conhecido na literatura a propor uma representação holística de cenários em um ambiente computacional foi proposto por Oliva e Torralba [11]. 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) em vez de analisá-lo apenas como um conjunto de objetos.

No final da década passada, o reconhecimento de cenários utilizando técnicas baseadas em levantamento de atributos tornou-se bastante popular, como notado por Zhou et al. em [12]. Uma das principais tentativas nesta direção foi proposta por Fei-Fei e Perona [13], 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. [14], 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 grupo 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 espacial 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 et al. [15]. 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 [16], quando as redes neurais convolucionais (“CNN” em inglês) 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 colocado<sup>1</sup> [18]. Desde então, os modelos de estado da arte em reconhecimento de objetos e cenários vêm utilizando as CNN.

Uma rede neural convolucional se diferencia de outras arquiteturas principalmente pela adição de camadas preparatórias que realizam operações denominadas de convolução. A operação de convolução divide a imagem em partes e as processa com o objetivo de

---

<sup>1</sup>Conforme [17], 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  $erro = \sum_{i=1}^N \sum_{j=1}^5 \min_j d_{ij}$ .

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 aplicar os mesmos coeficientes multiplicativos (ou pesos) em todas as partes da imagem.

A outra operação fundamental nas redes convolucionais é o *pooling*, uma operação de subamostragem que visa reduzir as dimensões das imagens, com o objetivo de torná-las resistentes a pequenas translações ou ruídos locais e também diminuindo o número de parâmetros a serem treinados.

As operações de convolução e pooling funcionam de forma análoga à seleção 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. Embora a seleção de atributos possa ser facilitada com o uso dessas redes, há um trabalho adicional na seleção dos chamados Hiperparâmetros. Bengio [19] define Hiperparâmetro da seguinte forma (traduzido de [19]): “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 pesos do modelo são ajustados para minimizar a função de custo a cada iteração do algoritmo de otimização, quantidade de camadas (*Number of Hidden Layers*) e quantidade de neurônios nas camadas intermediárias da rede (*Number of Hidden Units*).

O modelo AlexNet é uma rede neural convolucional de oito camadas: cinco convolucionais e três camadas completamente conectadas (ver Figura 2.1, extraída de [20]). A primeira camada convolucional processa uma imagem de dimensões 224 x 224 x 3 dividindo-a em 96 quadrados menores (chamados núcleos ou *kernels*) de dimensões 11 x 11 x 3 com passo (*stride*, em inglês) igual a 4. O passo é a distância entre os centros de núcleos vizinhos, que também pode ser entendido como a quantidade de pixels a ser deslocada por entre os núcleos dentro da imagem a cada processamento. A segunda camada convolucional recebe a saída da primeira após operações de normalização<sup>2</sup> e max-pooling<sup>3</sup> e aplica convolução com 256 núcleos de dimensões 5 x 5 x 48. A terceira camada convolucional recebe a saída da segunda novamente após operações de normalização e max-pooling e aplica convolução com 384 núcleos de dimensões 3 x 3 x 256. A terceira, quarta e quinta camadas aplicam mais convoluções e estão conectadas entre si sem operações de

---

<sup>2</sup>O tipo de normalização aplicada, chamado de *Lateral Response Normalization*, segue o método descrito na seção 3.3 do artigo original de Krizhevsky et al. [16].

<sup>3</sup>*Max-pooling* é uma operação de *pooling* que seleciona o valor máximo dos pixels de uma parte da imagem, descartando valores menores.

normalização e max-pooling. A saída da última camada convolucional serve de entrada para as camadas completamente conectadas, as duas primeiras com 4096 neurônios cada e a camada de saída com 1000 neurônios (um para cada classe de saída). A camada de saída aplica a função softmax, dada por  $\sigma(x_i) = \frac{e^{x_i}}{\sum_1^n e^{x_i}}$ , que pode ser entendido como uma generalização da função logística para um número de classes maior que 2.

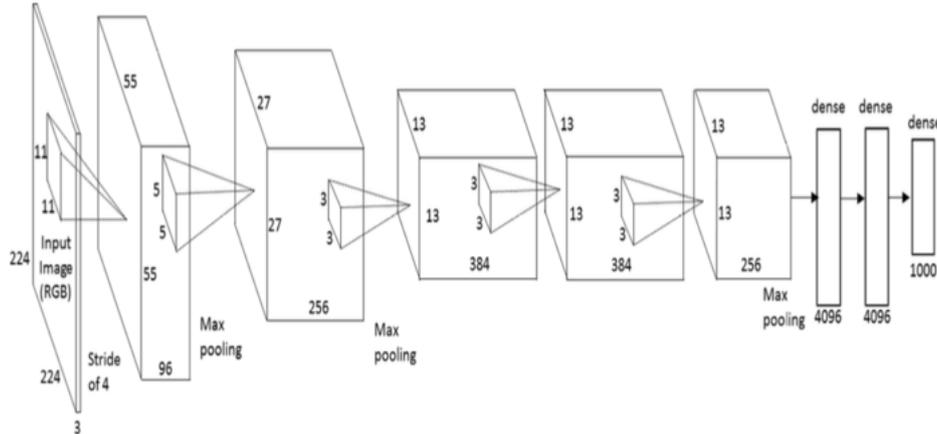


Figura 2.1: Evolução das dimensões da imagem através da AlexNet CNN.

O modelo AlexNet foi aperfeiçoado em trabalhos posteriores utilizando arquiteturas mais profundas, como nas redes VGG Net [21], GoogLeNet [22] e ResNet [23]. Entretanto, conforme resultados encontrados em [24], a diferença de performance entre a AlexNet e as arquiteturas mais modernas para classificação de cenários, se considerarmos a utilização de conjuntos de dados suficientemente grandes e diversos, ainda é pequena. Portanto, propomos a utilização da AlexNet para classificação de cenários como passo inicial do trabalho proposto.

## 2.3 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 modelos e algoritmos, mas também à disponibilidade de grandes quantidades de dados. De acordo com [24], 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 [25]. A rede neural vencedora desta competição em 2015, ResNet [23], 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 [26], MIT Indoor67 [27], Scene15 [15] e SUN Attribute [28]. Entretanto, o conjunto de dados de cenários com maior cobertura e densidade atualmente é o banco Places. Criado por Zhou et al [24], este banco é composto por mais de 10 milhões de imagens divididas entre 434 categorias de cenário.

Como ponto de partida deste trabalho, utilizaremos os atributos<sup>4</sup> aprendidos de uma rede de arquitetura AlexNet pré-treinada em um subconjunto do banco Places denominado Places365. O 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). Possui 1.803.460 imagens de treino, 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.

---

<sup>4</sup>Os atributos a serem extraídos, bem como de quais camadas esta extração será feita, serão definidos na etapa “Obter as classes de cenários” ver cronograma da seção 3.3.

# Capítulo 3

## Solução Proposta

Neste capítulo, abordamos as principais etapas a serem abordadas ao longo do projeto para atingirmos os objetivos geral e específicos apresentados na Seção 1.5. Algumas medidas de sucesso que poderão ser utilizadas para verificar o atendimento destes objetivos são:

- aumento na quantidade de fornecedores de fachada descobertos pela CGU e suas unidades regionais durante o período de um ano após a implantação deste trabalho;
- 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, é necessário o acesso a algumas bases de dados do Governo Federal, tais como:

- 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 do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE): pertencente ao Instituto Brasileiro de Geografia e Estatística (IBGE), contém uma lista de mais de 78 milhões de endereços urbanos e rurais distribuídos por mais de 300 mil setores censitários, classificados entre unidades residenciais, unidades de ensino, unidades de saúde e outros. Esta base, atualizada a partir do censo de 2010, pode ser utilizada para validação de diversas categorias de cenários obtidas através de uma rede convolucional.
- 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.

Para atingirmos os objetivos definidos, devem ser utilizadas ainda metas específicas para as atividades de Mineração de Dados, como as listadas a seguir:

- Medida da acurácia da classificação de cenários suspeitos, com ênfase em sensibilidade (taxa de verdadeiros positivos). Classificar um fornecedor não suspeito como suspeito possui um custo maior para a auditoria que classificar um fornecedor suspeito como não suspeito, uma vez que verificar se determinado fornecedor é ou não de fachada demanda um tempo precioso dos auditores do órgão;
- Medida da importância da classificação de cenários como componente de um modelo geral de classificação de fornecedores de fachada. A presença de um fornecedor em um cenário atípico é uma informação que deve aumentar a acurácia de um modelo mais geral utilizado para classificação de fornecedores de fachada, podendo ser medido pelo aumento do coeficiente de determinação  $R^2$  (em caso de um modelo final utilizando Regressão Logística) ou pela diminuição do erro quadrático médio

(para um modelo geral de classificação discreta, como Árvore de Decisão ou Redes Neurais).

## 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:

- realizar pesquisa bibliográfica;
- obter as imagens dos fornecedores;
- obter classes de cenários;
- realizar ajuste fino da CNN;
- realizar a validação e teste dos modelos;
- realizar implantação do modelo final.

Estas tarefas estarão distribuídas ao longo do Cronograma de Projeto exibido ao final deste Capítulo. Explicaremos cada uma das tarefas propostas nas próximas seções. A metodologia para atividades de mineração de dados foi baseada na metodologia do *Cross Industry Standard Process for Data Mining* (CRISP-DM) [29].

### 3.1.1 Realizar pesquisa bibliográfica

A primeira etapa do projeto é realizar um levantamento bibliográfico sobre Modelos de Detecção de Fraudes (com foco em Detecção de Fornecedores de Fachada) e sobre Modelos de Reconhecimento de Cenários, desde os mais tradicionais como BoVW e SPM até os modelos baseados em CNN referentes ao estado da arte. Esta etapa já foi realizada neste trabalho e está descrita no Capítulo 2 (Estado da arte).

### 3.1.2 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 Governo Federal, o que, na maior parte dos casos<sup>1</sup>, é o suficiente para acessarmos as imagens de suas sedes através do Google Street View.

Entretanto, há casos em que os endereços obtidos ainda 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. Após a requisição pelos metadados, a API retorna status de resposta “OK” caso seja possível recuperar a imagem com sucesso<sup>2</sup>.

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 pode 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 AlexNet pré-treinada a partir do conjunto de cenários Places [24], que, a exemplo do Street View, é composta apenas de imagens em nível de rua.

Entretanto, a classificação de imagens de satélite do Google Maps pode servir de 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.

### 3.1.3 Obter classes de cenários

Como será visto na Seção 3.2 (Experimentos Realizados), um dos principais problemas encontrados ao utilizarmos uma rede AlexNet pré-treinada para classificação de cenários de fornecedores foi que as classificações encontradas podem não ser adequadas aos cenários brasileiros. Para atingirmos resultados mais efetivos, propomos aplicar técnicas de Clusterização (Hierárquica ou através do algoritmo K-Means) a partir dos atributos gerados pela AlexNet, a fim de obtermos grupos de cenários bem definidos a partir das imagens obtidas.

Ao aplicarmos técnicas de clusterização nos atributos gerados de uma rede AlexNet pré-treinada, espera-se encontrar grupos de cenários principais. A partir da amostragem

---

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

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

desses grupos e com a ajuda de auditores da CGU especialistas no domínio “Fornecedores de Fachada”, será possível rotular cada um dos grupos identificados em classes que reflitam de forma mais adequada o contexto dos cenários brasileiros. Além disso, será possível aos auditores definir, dentre os grupos existentes, aqueles que representam um maior grau de atipicidade ou suspeição em relação às atividades exercidas por estes fornecedores.

A partir das classes obtidas através da aplicação de técnicas de clusterização e ajuda de especialistas do domínio, deve-se validar os conjuntos obtidos, de forma a garantir que nenhuma imagem tenha sido rotulada em uma classe equivocadamente. Um erro de classificação nestes conjuntos de imagens pode afetar a acurácia e a sensibilidade do modelo final.

### 3.1.4 Realizar ajuste fino da CNN

Uma vez gerado um conjunto de imagens com classes adequadas ao contexto brasileiro, podemos refinar a arquitetura AlexNet para identificar as novas classes. Realizaremos este refinamento de três modos, dando origem a três modelos básicos distintos:

- Modelo 1: Obtido pelo treinamento de uma rede AlexNet a partir do início, sem utilização de pesos pré-treinados;
- Modelo 2: Obtido pelo treinamento de uma rede AlexNet, porém utilizando pesos iniciais já aprendidos a partir do dataset Places365 [24];
- Modelo 3: Obtido pelo treinamento de um classificador SVM a partir dos atributos da rede Alexnet pré-treinada a partir do dataset Places365 [24].

Os dois últimos modelos propostos ilustram técnicas conhecidas de Transferência de Aprendizagem (*Transfer Learning* em inglês). Conforme [30], a Transferência de Aprendizagem 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.

A partir dos modelos base listados, geraremos modelos derivados através da substituição da rede AlexNet por arquiteturas mais modernas: GoogLeNet e ResNet<sup>3</sup>. Portanto, teremos gerado nove modelos de CNN, equivalentes aos três tipos de arquitetura para os três modelos básicos listados.

---

<sup>3</sup>Embora a rede VGG Net também seja uma CNN mais profunda que a AlexNet, optamos pela utilização das redes GoogLeNet e ResNet, por serem mais atuais.

### 3.1.5 Realizar a validação e teste dos modelos

As abordagens levantadas na seção anterior serão avaliadas através de um conjunto de imagens de validação. A métrica desejada deve privilegiar a precisão e não a sensibilidade<sup>4</sup>

Neste trabalho, evitamos o uso do *F-score*<sup>5</sup> por entendermos que a característica que deve ser maximizada é a precisão. Ou seja, queremos que os fornecedores classificados como atípicos sejam de fato de interesse dos auditores, dado que um fornecedor identificado como suspeito demandará custos de verificação por meio de inspeções ou auditorias, e, embora a proporção com que classificamos um fornecedor suspeito como não suspeito também seja relevante, esta informação não traz custos adicionais para a auditoria.

Os resultados das métricas de validação poderão ser utilizados para refinar os hiperparâmetros (ver Seção 2.2) dos modelos testados.

Por fim, os resultados das métricas selecionadas no conjunto de teste serão utilizados para selecionarmos a melhor abordagem de modelo de rede dentre os modelos treinados na Seção 3.1.4.

### 3.1.6 Realizar implantação do modelo final

Como visto na Seção 2.1 de nossa revisão de Estado da Arte, os modelos de detecção de fornecedores de fachada, especialmente aqueles que buscam detectar declarações financeiras falsas, utilizam em sua maioria técnicas como Regressão Logística, Árvore de Decisão e Redes Neurais. Estes modelos geralmente se baseiam em informações financeiras das empresas, tais como balanços patrimoniais e demonstrações contábeis.

O modelo final a ser implantado no âmbito da CGU será uma combinação destes modelos tradicionais, utilizando informações já levantadas pela Controladoria, tais como: quantidade e nível de escolaridade de empregados da empresa (a partir da base da RAIS), quantidade e qualificação dos sócios (a partir da base da RFB), valores de contratos prévios com o Governo Federal (a partir da base SIASG Comprasnet) e classificação censitária do endereço da empresa (a partir da base do CNEFE), com as informações de classes de cenários obtidas através da aplicação de CNN.

Conforme as hipóteses elencadas para este trabalho na Seção 1.4, espera-se que a inclusão de informações de classes de cenários possa auxiliar na melhoria da acurácia e da sensibilidade do modelo final, que será integrado ao modelo de ranqueamento de contratos

---

<sup>4</sup>A métrica de precisão calcula a proporção de verdadeiros positivos para cada classificação retornada pelo modelo. É calculado pela fórmula  $\frac{\sum \text{Verdadeiros Positivos}}{\sum \text{Classificados como positivos}}$ . A sensibilidade calcula a taxa de verdadeiros positivos para cada classe do conjunto de dados. É dada pela fórmula  $\frac{\sum \text{Verdadeiros Positivos}}{\sum \text{Elementos da classe}}$ .

<sup>5</sup>O *F-score*, ou Medida F, busca equilibrar as métricas de precisão e sensibilidade, sendo dado pela fórmula  $F_1 = \frac{2 \times \text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$ .

públicos [3] e ao atual Sistema APC da CGU, que poderão ser consultados posteriormente pelos auditores do órgão.

## 3.2 Experimentos Realizados

Como experimento inicial para comprovar a viabilidade da classificação de cenários de fornecedores do Governo Federal, foram classificados cerca de 15 mil cenários de fornecedores<sup>6</sup> através de uma rede AlexNet pré-treinada a partir do conjunto de cenários Places205<sup>7</sup>. A rede AlexNet utilizada foi implementada para o framework de *deep learning* Caffe<sup>8</sup>, e as imagens foram classificadas usando a linguagem Python. Imagens mais recentes dos últimos 12 meses serão classificadas conforme a evolução deste projeto. Os cenários mais frequentes encontrados pelo modelo estão na Figura 3.1.

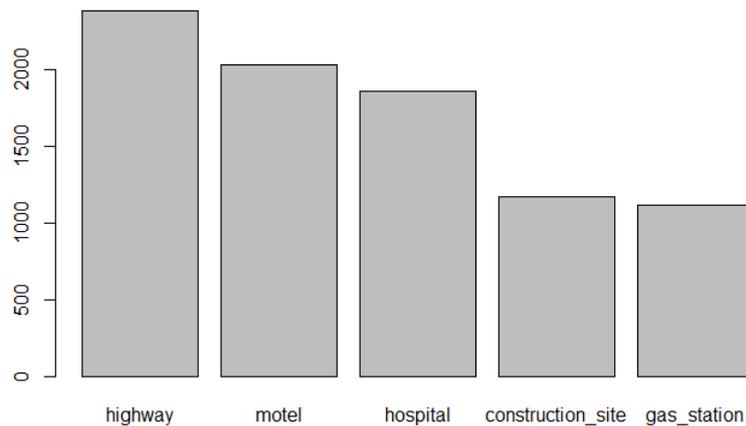


Figura 3.1: As cinco categorias de cenários mais comuns encontradas pela rede AlexNet.

Através do levantamento de uma amostra aleatória simples de tamanho  $n = 375$  (conforme metodologia adotada em Cochran [33]), encontramos as acurácias top 1 e top 5

---

<sup>6</sup>Obtidos através de endereços de fornecedores da base da RFB que possuíam contratos ativos com o Governo Federal até 28 de setembro de 2016 pela base do sistema SIASG Comprasnet. As imagens, baixadas através da API do Google Street View, possuem tamanho 600x400 e resolução de 96 pontos por polegada.

<sup>7</sup>O Places205 é uma versão anterior do conjunto Places365, que pretendemos utilizar ao longo deste projeto. De acordo com [31], este conjunto possui cerca de 7 milhões de imagens classificadas em mais de 200 categorias de cenários.

<sup>8</sup>Caffe é um framework de *deep learning* criado pelo laboratório de Inteligência Artificial da Universidade da Califórnia em Berkeley (UC Berkeley), disponível em <http://caffe.berkeleyvision.org/>. Contém uma série de algoritmos de estado da arte e modelos de referência pré-treinados [32]. O modelo AlexNet pré-treinado encontra-se disponível em <http://places.csail.mit.edu/>.

do modelo (top 1 refere-se à classe mais provável descoberta pela rede e top 5 refere-se a uma das cinco classes mais prováveis descobertas pela rede). A acurácia top 1 encontrada foi de 33,6%, enquanto a acurácia top 5 do modelo foi de 83,7% (ambas calculadas com erro permissível de 5% e nível de confiança de 95%).

O fato da acurácia top 5 do modelo ser muito superior à acurácia top 1 nos mostra que, na maior parte das vezes, o modelo identifica classes próximas daquela que descreve corretamente a imagem. Como estamos tratando de cenários, um rótulo apenas pode ser insuficiente para descrever uma imagem ou uma mesma imagem pode ser descrita por vários rótulos. Isto nos sugere que utilizemos uma classificação multilabel no decorrer deste Projeto.

Uma ressalva feita ao resultado encontrado refere-se ao excesso de cenários classificados como “highway” (estradas). Isto provavelmente se deve ao fato de que, no contexto do Google Street View, as fotos geralmente mostram as ruas por padrão, já que são nelas que normalmente trafegam os veículos da empresa. Ao longo da execução do projeto, serão baixadas mais de uma imagem por fornecedor, com mudança nos parâmetros de orientação da imagem (“heading”), para evitarmos este tipo de problema.

Como o padrão de estrada que predomina nos cenários coletados são simples de serem detectados, o resultado da acurácia obtida está superestimado. Se excluirmos de nossa amostra os resultados classificados como estrada, obtemos as seguintes acurácias: 23,6% (top 1) e 75,9% (top 5).

Um resultado parcial de interesse diz respeito a um grupo de 103 imagens classificadas como “slum” (favelas). Muitos dos cenários classificados nesta categoria não seriam considerados favelas no contexto brasileiro. Nos resultados obtidos da classificação das imagens da Figura 3.2, a primeira coluna (A) mostra cenários corretamente classificados como “loja de roupas” e “floresta”, enquanto a segunda coluna (B) mostra resultados classificados incorretamente como “prisão” e “favela”. Os erros de classificação da coluna B podem ter sido causados pela presença de barras na primeira imagem e pela presença de cabos elétricos na segunda (muito comum em bairros residenciais brasileiros).

Estes resultados indicam de fato a necessidade de criarmos classes próprias para o projeto, considerando os cenários brasileiros para este propósito. A validação destas novas classes, conforme definido na seção 3.1.3 de nossa metodologia, ficará a cargo dos próprios auditores da CGU.

### **3.3 Cronograma**

As tarefas a serem desempenhadas para atingir os objetivos gerais e específicos deste Projeto dividem-se de acordo com o calendário da Tabela 3.1.



Figura 3.2: Cenários classificados pela rede AlexNet.

Tabela 3.1: Cronograma

Tarefas	ago/17	set/17	out/17	nov/17	dez/17	jan/18	fev/18	mar/18	abr/18	mai/18	jun/18	jul/18
Realizar pesquisa bibliográfica	X											
Obter as imagens dos fornecedores		X										
Obter classes de cenários			X	X								
Realizar ajuste fino da CNN					X	X						
Realizar a validação e teste dos modelos							X	X				
Realizar implantação do modelo final									X	X		
Escrever Dissertação											X	
Defesa da Dissertação e Publicação do Trabalho												X

# Capítulo 4

## Resultados Esperados

Embora não seja razoável esperar que apenas a classificação de cenários seja suficiente para a identificação de fornecedores de fachada, espera-se que sirva de variável relevante para aperfeiçoar modelos preditivos com mais atributos como, por exemplo, o Modelo de Riscos de Contratos Públicos utilizado na CGU [3], bem como aperfeiçoar sistemas internos, como o Sistema APC.

A utilização de imagens como uma fonte de dados também representa uma inovação nos trabalhos da Controladoria, dado que as informações utilizadas pelos auditores provém principalmente de fontes estruturadas, como bancos de dados de sistemas governamentais. A possibilidade de se trabalhar com imagens certamente abrirá novos horizontes para futuras aplicações que sirvam de suporte ao Sistema de Controle Interno do Governo Federal.

# Referências

- [1] Dassen, Nicolás e Juan Cruz Vieyra: *Open Government and Targeted Transparency*. Trends and Challenges for Latin America and the Caribbean, Washington, DC, Banco Interamericano de Desarrollo, 2012. [http://www.academia.edu/download/41609023/Open\\_Government\\_and\\_Targeted\\_Transparency\\_\\_Trends\\_and\\_Challenges\\_for\\_Latin\\_America\\_and\\_the\\_Caribbean.pdf](http://www.academia.edu/download/41609023/Open_Government_and_Targeted_Transparency__Trends_and_Challenges_for_Latin_America_and_the_Caribbean.pdf), acesso em 2017-08-02. 1
- [2] Willebois, Emile van der Does de, J. C. Sharman, Robert Harrison, Ji Won Park e Emily Halter: *The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It*. World Bank Publications, novembro 2011, ISBN 978-0-8213-8896-9. Google-Books-ID: WdTJ6LPhBxYC. 2, 6
- [3] Sales, Leonardo Jorge: *Proposta de modelo de classificação do risco de contratos públicos*. Tese de Mestrado, Universidade de Brasília, Faculdade de Economia, Administração e Contabilidade – FACE/UnB - Universidade de Brasília - Campus Darcy Ribeiro - Prédio da FACE - Asa Norte - CEP: 70910-900 - Brasília - DF, setembro 2016. Trabalho apresentado como requisito parcial para obtenção do título de Mestre em Economia do Setor Público. 3, 4, 5, 8, 19, 22
- [4] Fawcett, Tom e Foster Provost: *Adaptive fraud detection*. Data mining and knowledge discovery, 1(3):291–316, 1997. 6
- [5] Neville, Jennifer e David Jensen: *Iterative classification in relational data*. Em *Proc. AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, páginas 13–20, 2000. 7
- [6] Bolton, Richard J e David J Hand: *Statistical fraud detection: A review*. Statistical science, páginas 235–249, 2002. 7
- [7] Albashrawi, Mousa: *Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015*. Journal of Data Science, 14(3):553–569, 2016. 7, 8
- [8] Spathis, Charalambos T: *Detecting false financial statements using published data: some evidence from greece*. Managerial Auditing Journal, 17(4):179–191, 2002. 7
- [9] Griffith, Arnold K: *Mathematical models for automatic line detection*. Journal of the ACM (JACM), 20(1):62–80, 1973. 8
- [10] Griffith, Arnold K: *Edge detection in simple scenes using a priori information*. IEEE Transactions on Computers, 100(4):371–381, 1973. 8

- [11] Oliva, Aude e Antonio Torralba: *Modeling the shape of the scene: A holistic representation of the spatial envelope*. International journal of computer vision, 42(3):145–175, 2001. 8
- [12] Zou, Jinyi, Wei Li, Chen Chen e Qian Du: *Scene classification using local and global features with collaborative representation fusion*. Information Sciences, 348:209–226, 2016. 9
- [13] Fei-Fei, Li e Pietro Perona: *A bayesian hierarchical model for learning natural scene categories*. Em *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, páginas 524–531. IEEE, 2005. 9
- [14] Yang, Jun, Yu Gang Jiang, Alexander G Hauptmann e Chong Wah Ngo: *Evaluating bag-of-visual-words representations in scene classification*. Em *Proceedings of the international workshop on Workshop on multimedia information retrieval*, páginas 197–206. ACM, 2007. 9
- [15] Lazebnik, Svetlana, Cordelia Schmid e Jean Ponce: *Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories*. Em *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, volume 2, páginas 2169–2178. IEEE, 2006. 9, 12
- [16] Krizhevsky, Alex, Ilya Sutskever e Geoffrey E Hinton: *Imagenet classification with deep convolutional neural networks*. Em *Advances in neural information processing systems*, páginas 1097–1105, 2012. 9, 10
- [17] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg e Li Fei-Fei: *ImageNet Large Scale Visual Recognition Challenge*. arXiv:1409.0575 [cs], setembro 2014. <http://arxiv.org/abs/1409.0575>, acesso em 2017-12-22, arXiv: 1409.0575. 9
- [18] Harada, Tatsuya e Yasuo Kuniyoshi: *Graphical gaussian vector for image categorization*. Em *Advances in Neural Information Processing Systems*, páginas 1547–1555, 2012. 9
- [19] Bengio, Yoshua: *Practical recommendations for gradient-based training of deep architectures*. Em *Neural networks: Tricks of the trade*, páginas 437–478. Springer, 2012. 10
- [20] Shoieb, Doaa A, Sherin M Youssef e Walid M Aly: *Computer-aided model for skin diagnosis using deep learning*. Journal of Image and Graphics, 4(2), 2016. 10
- [21] Simonyan, Karen e Andrew Zisserman: *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv:1409.1556 [cs], setembro 2014. <http://arxiv.org/abs/1409.1556>, arXiv: 1409.1556. 11
- [22] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke e Andrew Rabinovich: *Going deeper with convolutions*. Em *Proceedings of the IEEE*

- conference on computer vision and pattern recognition*, páginas 1–9, 2015. [http://www.cv-foundation.org/openaccess/content\\_cvpr\\_2015/html/Szegedy\\_Going\\_Deeper\\_With\\_2015\\_CVPR\\_paper.html](http://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html), acesso em 2017-08-20. 11
- [23] He, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun: *Deep Residual Learning for Image Recognition*. arXiv:1512.03385 [cs], dezembro 2015. <http://arxiv.org/abs/1512.03385>, arXiv: 1512.03385. 11, 12
- [24] Zhou, Bolei, Aditya Khosla, Agata Lapedriza, Antonio Torralba e Aude Oliva: *Places: An image database for deep scene understanding*. arXiv preprint arXiv:1610.02055, 2016. 11, 12, 16, 17
- [25] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg e Li Fei-Fei: *ImageNet Large Scale Visual Recognition Challenge*. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 12
- [26] Xiao, Jianxiong, James Hays, Krista A Ehinger, Aude Oliva e Antonio Torralba: *Sun database: Large-scale scene recognition from abbey to zoo*. Em *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, páginas 3485–3492. IEEE, 2010. 12
- [27] Quattoni, Ariadna e Antonio Torralba: *Recognizing indoor scenes*. Em *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, páginas 413–420. IEEE, 2009. 12
- [28] Patterson, Genevieve e James Hays: *Sun attribute database: Discovering, annotating, and recognizing scene attributes*. Em *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, páginas 2751–2758. IEEE, 2012. 12
- [29] Wirth, Rüdiger e Jochen Hipp: *Crisp-dm: Towards a standard process model for data mining*. Em *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, páginas 29–39, 2000. 15
- [30] Yosinski, Jason, Jeff Clune, Yoshua Bengio e Hod Lipson: *How transferable are features in deep neural networks?* Em *Advances in neural information processing systems*, páginas 3320–3328, 2014. 17
- [31] Zhou, Bolei, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba e Aude Oliva: *Learning deep features for scene recognition using places database*. Em *Advances in neural information processing systems*, páginas 487–495, 2014. 19
- [32] Jia, Yangqing, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama e Trevor Darrell: *Caffe: Convolutional architecture for fast feature embedding*. arXiv preprint arXiv:1408.5093, 2014. 19
- [33] Cochran, W.G.: *Técnicas de amostragem*. Ed. Fundo de Cultura, 1965. <https://books.google.com.br/books?id=dz2AGwAACAAJ>. 19