

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE COMPUTAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**GUILHERME RIGHETTO**

**O USO DA REDE NEURAL CONVOLUCIONAL  
COMO EXTRATOR DE CARACTERÍSTICAS  
APLICADO AO PROBLEMA DE IDENTIFICAÇÃO  
DE ESCRITORES**

MONOGRAFIA

**CAMPO MOURÃO**

**2016**

**GUILHERME RIGHETTO**

**O USO DA REDE NEURAL CONVOLUCIONAL  
COMO EXTRATOR DE CARACTERÍSTICAS  
APLICADO AO PROBLEMA DE IDENTIFICAÇÃO  
DE ESCRITORES**

Trabalho de Conclusão de Curso de Graduação apresentado à disciplina de Trabalho de Conclusão de Curso 2, do Curso de Bacharelado em Ciência da Computação do Departamento Acadêmico de Computação da Universidade Tecnológica Federal do Paraná, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Dr. Diego Bertolini Gonçalves

**CAMPO MOURÃO**

**2016**



## ATA DA DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

Às dezoito horas e trinta minutos do dia vinte e dois de novembro de dois mil e dezesseis foi realizada no Miniauditório de EAD a sessão pública da defesa do Trabalho de Conclusão do Curso de Bacharelado em Ciência da Computação do acadêmico **Guilherme Righetto** com o título **O uso da rede neural convolucional como extrator de características aplicado ao problema de identificação de escritores**. Estavam presentes, além do acadêmico, os membros da banca examinadora composta pelo professor Dr. Diego Bertolini Gonçalves (Orientador-Presidente), pela professora Dr<sup>a</sup>. Aretha Barbosa Alencar e pelo professor Dr. Frank Helbert Borsato. Inicialmente, o aluno fez a apresentação do seu trabalho, sendo, em seguida, arguido pela banca examinadora. Após as arguições, sem a presença do acadêmico, a banca examinadora o considerou **aprovado** na disciplina de Trabalho de Conclusão de Curso e atribuiu, em consenso, a **nota 9,0 (nove)**. Este resultado foi comunicado ao acadêmico e aos presentes na sessão pública. A banca examinadora também comunicou ao acadêmico que este resultado fica condicionado à entrega da versão final dentro dos padrões e da documentação exigida pela UTFPR ao professor Responsável do TCC no prazo de onze dias. Em seguida foi encerrada a sessão e, para constar, foi lavrada a presente Ata que segue assinada pelos membros da banca examinadora, após lida e considerada conforme.

Observações:

---

---

---

---

Campo Mourão, 22 de novembro de 2016

Prof<sup>a</sup>. Dr<sup>a</sup>. Aretha Barbosa Alencar  
Membro

Prof. Dr. Frank Helbert Borsato  
Membro

Prof. Dr. Diego Bertolini Gonçalves  
Orientador

A ata de defesa assinada encontra-se na coordenação do curso.

# Agradecimentos

---

Agradeço à minha família, em especial meus pais João e Iracema. Por sempre me incentivarem ao decorrer do curso. E por estarem sempre perto, mesmo com a distância.

Aos amigos que fiz durante toda a graduação, na qual se tornaram minha segunda família, em especial a todos que fizeram parte da República Molejão de alguma forma.

Ao meu orientador, Diego Bertolini Gonçalves por me proporcionar conhecimento e realizar diversas correções e discussões para o desenvolvimento deste trabalho.

Aos professores da banca avaliadora, Aretha Barbosa Alencar, Juliano Foleiss e Frank Helbert Borsato, por suas contribuições a respeito do trabalho.

Aos demais professores por me propiciarem conhecimentos fundamentais em computação e sempre estarem dispostos para ajudar sobre qualquer dúvida.

Muito obrigado!

# Resumo

---

Righetto, Guilherme. O uso da rede neural convolucional como extrator de características aplicado ao problema de identificação de escritores. 2016. 49. f. Monografia (Curso de Bacharelado em Ciência da Computação), Universidade Tecnológica Federal do Paraná. Campo Mourão, 2016.

**Contexto:** No contexto da identificação de escritores, frequentemente pesquisadores propõem diferentes métodos para a extração, processamento das características e classificação. Podemos dividir a identificação de escritores em dois grandes grupos. O primeiro extrai características locais referentes a escrita, como espaçamento, vales, laços, angulações, dentre outras. O segundo representa a escrita através de descritores de textura, que extraem características globais. As bases de documentos manuscritos normalmente possuem um único estilo de escrita. Entretanto atualmente o conteúdo de cada amostra do mesmo escritor produz documentos em línguas diferentes, como, português, árabe, inglês, alemão, dentre outras. Ao considerar mais de um estilo de escrita para o problema de identificação de escritores, a tarefa torna-se ainda mais difícil, pois é necessário construir um sistema de identificação independente do alfabeto utilizado. A fim de solucionar o problema de identificação de escritores que produzem documentos em línguas diferentes, foram utilizadas diversas técnicas como citado anteriormente. Uma delas é utilizar uma rede neural convolucional (CNN) como extrator de características e classificador, na qual, também foi utilizada neste trabalho, além da abordagem de dissimilaridade, que transforma um problema *n-classes* em binário.

**Objetivo:** O principal objetivo deste trabalho é avaliar o desempenho propiciado através das características extraídas pela rede neural convolucional (CNN) no processo de identificação *off-line* de escritores. Para isso utilizaremos as bases de dados BFL, CVL e QUWI.

**Método:** O método proposto neste trabalho cumpriu as seguintes etapas: Realizou um pré-processamento nos documentos manuscritos utilizando uma abordagem de geração de textura. Posteriormente a textura foi dividida em blocos de tamanho igual a  $128 \times 128$ . Na etapa seguinte com uso do classificador CNN / extrator universal de características foi utilizado dois recursos providos pela CNN, sendo a classificação de cada bloco de entrada. Portanto, o processo tradicional do classificador CNN e também o uso da CNN como um extrator de características. Em seguida, o cálculo da dissimilaridade sobre os vetores de características extraídos a partir de cada bloco. Na etapa seguinte foi utilizado o classificador

SVM para classificar os blocos de textura. Por fim, foi realizado a combinação das predições geradas de cada bloco, pelo SVM e pela CNN, a fim de obter uma decisão final sobre quem escreveu determinado documento.

**Resultados:** Os principais resultados obtidos foram utilizando a abordagem de dissimilaridade no vetor de características extraído pela rede neural convolucional, sendo eles 98,26% (BFL), 97,91% (CVL) e 86,96% (QUWI).

**Conclusões:** Concluimos neste trabalho, que a abordagem de dissimilaridade se mantém robusta em relação a identificação de escritores de documentos manuscritos de diferentes línguas. Além disso, observou-se que as características extraídas pela CNN obteve bons resultados nos casos em que a língua escrita seguia um alfabeto similar.

**Palavras-chaves:** Rede Neural Convolucional, Identificação de escritores, Dissimilaridade.

# Abstract

---

Righetto, Guilherme. . 2016. 49. f. Monograph (Undergraduate Program in Computer Science), Federal University of Technology – Paraná. Campo Mourão, PR, Brazil, 2016.

**Context:** In the context of the writer identification, researchers often propose different methods for extraction, processing of features and classification. In which we can divide the writer identification in two groups. The first one extracted local features related to writing, like as spacing, concavity, angulations, among others. The second represents writing through texture descriptors, which extract global features. Bases of handwritten documents usually have a unique writing style. However, currently the content of each sample of the same writer produces documents in different languages, such as Portuguese, Arabic, English, German, among others. When considering more than one writing style for the problem of identifying writers, the task becomes even more difficult as it is necessary to construct an identification system independent of the alphabet used. In order to solve the problem of writer identification who produce documents in different languages, several techniques were used as mentioned above. One of them is to use a convolutional neural network (CNN) as a feature extractor and classifier, in which it was also used in this work, in addition to the dissimilarity approach, which turns a *n-classes* problem into binary.

**Objective:** The main objective of this work is to evaluate the performance provided by the features extracted by the convolutional neural network (CNN) in the process of the off-line writer identification. For this we will use the databases BFL, CVL and QUWI.

**Method:** The method proposed in this work fulfilled the following steps: Preprocessed manuscript documents using a texture generation approach. Subsequently the texture was divided into blocks of different sizes. In the next step using the CNN classifier / universal features extractor, two CNN resources were used, the classification of each input block, that is, the traditional CNN classifier process and also the use of CNN as a features extractor. Then, dissimilarity feature vectors are computed by feature vectors extracted from each block. In the next step, the SVM classifier was used to classify the texture blocks. Finally, a combination of the predictions generated from each block was made by the SVM and CNN in order to obtain a final decision on who wrote a particular document.

**Results:** The main results were obtained using the dissimilarity approach in the feature vector extracted by the convolutional neural network, 98.26% (BFL), 97.91% (CVL) and

86.96% (QUWI).

**Conclusions:** We conclude in this work that the dissimilarity approach remains robust in relation to the writer identification of handwritten documents of different languages. In addition, it was observed that the features extracted by CNN obtained good results in cases where the written language followed a similar alphabet.

**Keywords:** Convolutional Neural Network, Writer Identification, Dissimilarity



# Lista de figuras

---

3.1	Documento da base BFL. . . . .	21
3.2	Documento em árabe com escrita livre da base QUWI. . . . .	22
3.3	Documento em inglês da base CVL. . . . .	22
3.4	Geração de textura. (a) Alinhamento dos componentes e (b) a textura. . . . .	23
3.5	Exemplo de arquitetura da CNN. . . . .	24
3.6	Exemplo do campo receptivo local. . . . .	25
3.7	Exemplos de filtros da camada convolucional. . . . .	26
3.8	Exemplo de <i>max-pooling</i> . . . . .	27
3.9	Exemplo de dissimilaridade utilizando uma base de dados sintética. . . . .	28
3.10	Dissimilaridade (a) entre amostras do mesmo escritor para gerar amostras positivas e (b) entre amostras do escritor da classe positiva e diferentes escritores para gerar amostras negativas. . . . .	29
4.1	Visão geral do método proposto. . . . .	31
4.2	Bloco de tamanho $128 \times 128$ . . . . .	32
5.1	Experimentos realizados com a base BFL para analisar o desempenho em relação ao tamanho da penúltima camada. . . . .	36
5.2	Experimentos realizados com a base BFL para analisar o desempenho em relação ao tamanho do bloco. . . . .	37
5.3	Experimentos realizados com a base BFL para analisar o desempenho em relação a quantidade de blocos. . . . .	38
5.4	Amostra (a) escrita em inglês e amostra (b) escrita em alemão do mesmo escritor da base CVL. . . . .	39
5.5	Amostra (a) escrita em inglês e amostra (b) escrita em árabe do mesmo escritor da base QUWI. . . . .	40
5.6	Distribuição das amostras em inglês (1,2,4) e alemão (5) utilizando 115 escritores da base CVL. . . . .	42
5.7	Distribuição das amostras em árabe (1,2) e inglês (4) utilizando 115 escritores da base QUWI. . . . .	43

# Lista de tabelas

---

2.1	Resumo do estado da arte. . . . .	19
5.1	Configuração da arquitetura utilizada pela CNN. . . . .	36
5.2	Melhores Taxas de Identificação utilizando a Base BFL. . . . .	39
5.3	Taxa de Identificação utilizando a base CVL. . . . .	41
5.4	Taxa de Identificação utilizando a base QUWI. . . . .	42
5.5	Taxa de identificação utilizando outros 115 escritores de cada base. . . . .	44
5.6	Taxa de identificação utilizando todos escritores de cada base. . . . .	44

# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>11</b>
1.1	Motivação . . . . .	13
1.2	Desafios . . . . .	13
1.3	Objetivos . . . . .	14
1.4	Estrutura do trabalho . . . . .	14
<b>2</b>	<b>Estado da Arte</b>	<b>15</b>
2.1	Abordagens Tradicionais de Classificação . . . . .	15
2.2	Abordagens utilizando Aprendizagem Profunda . . . . .	17
2.3	Discussão . . . . .	18
<b>3</b>	<b>Fundamentação Teórica</b>	<b>20</b>
3.1	Base de Dados . . . . .	20
3.1.1	BFL . . . . .	20
3.1.2	QUWI . . . . .	21
3.1.3	CVL . . . . .	22
3.2	Geração de Textura . . . . .	22
3.3	Aprendizagem Profunda . . . . .	23
3.3.1	Rede Neural Convolutacional (CNN) . . . . .	24
3.3.2	Camada Convolutacional . . . . .	25
3.3.3	Camada de Pooling . . . . .	26
3.3.4	Camada Totalmente Conectada . . . . .	27
3.4	Dissimilaridade . . . . .	27
3.5	Combinação de Classificadores . . . . .	29
3.6	Considerações . . . . .	30
<b>4</b>	<b>Método</b>	<b>31</b>
4.1	Pré-Processamento . . . . .	32
4.2	Rede Neural Convolutacional . . . . .	32
4.2.1	Classificador . . . . .	33
4.2.2	Extrator Universal de Características . . . . .	33
4.3	Dissimilaridade . . . . .	33

4.4	Máquina de Vetores de Suporte (SVM) . . . . .	34
4.5	Combinação das predições . . . . .	34
4.6	Considerações . . . . .	34
<b>5</b>	<b>Experimentos</b>	<b>35</b>
<b>6</b>	<b>Conclusão</b>	<b>45</b>
6.1	Trabalhos Futuros . . . . .	45
	<b>Referências</b>	<b>47</b>

---

## Introdução

---

Segundo Jain et al. (2000), o reconhecimento de padrões é o estudo de como máquinas podem observar o ambiente, aprender a distinguir padrões e tomar decisões sobre as categorias dos padrões. Desta forma, podemos dizer que a importância do reconhecimento de padrões está diretamente ligada a encontrar métodos que automatizem tarefas, que possam ser exaustivas para os humanos, como também sujeitas a falhas. A fim de tentar solucionar este problema, pesquisadores propõem métodos para diversas tarefas de reconhecimento, tais como, faces (AHONEN et al., 2006), objetos (KRIZHEVSKY et al., 2012), escrita (DJEDDI et al., 2013), dentre outros. Neste trabalho utilizaremos a escrita como objeto de estudo a qual será utilizada para identificar o autor de determinado documento.

Ao trabalharmos com escrita em reconhecimento de padrões, podemos ter diferentes áreas de estudo. *Optical Character Recognition* é uma das áreas mais antigas a qual tem por objetivo realizar o reconhecimento de caracteres a partir de imagens. Segundo Trier et al. (1996) é uma das áreas mais bem sucedidas de reconhecimento de padrões. Existem diversos tipos de problemas na qual o reconhecimento de caracteres pode ser empregado, os principais são autenticação de assinaturas, o reconhecimento de estilos de escrita (idioma escrito), identificação de escritores, dentre outros. O foco de estudo neste trabalho é a identificação de escritores a partir de documentos manuscritos.

Os termos identificação e verificação são frequentemente citados na área de reconhecimento de padrões, entretanto existe uma grande diferença entre os dois. A verificação tem como objetivo avaliar se uma amostra pertence a uma determinada classe ou não. Enquanto a identificação avalia dentre um conjunto de classes candidatas, a quem pertence uma determinada instância (AMARAL, 2014).

Estudos com identificação de escritores utilizando documentos manuscritos podem utilizar documentos *online* ou *off-line*. Na técnica *online* são utilizados dispositivos capazes de registrar características temporais da escrita, por exemplo uma mesa digitalizadora. Nestes

casos existe menor número de ruídos, e diferentes características podem ser obtidas durante a escrita, apresentando desempenho superior aos métodos *off-line*. Enquanto na técnica *off-line*, as características são extraídas a partir do texto digitalizado. Aumentando o risco de apresentar ruídos durante a digitalização além de uma maior dificuldade em extrair características, apresentando-se como mais desafiadora.

Com relação ao texto, avaliaremos dois tipos de documentos manuscritos, sendo eles texto-dependente e texto-independente. Os documentos com texto-dependente, possuem um conteúdo predeterminado, ou seja, o conteúdo será o mesmo em todas as amostras. Enquanto, os documentos com texto-independente possuem o conteúdo diferente para cada escritor, ou seja, cada pessoa escreve um texto livre, existindo uma variação grande na quantidade de palavras.

No contexto da identificação de escritores, frequentemente pesquisadores propõem diferentes métodos para a extração, processamento das características e classificação. A etapa de extração de características tem impacto direto nas etapas seguintes. Se características extraídas não forem suficientemente boas para descrever determinada classe, o processo todo pode falhar.

Em geral podemos destacar dois grandes grupos das quais os métodos de extração podem se enquadrar. O primeiro seria baseado em abordagens forenses, extraíndo características locais referentes a escrita (AL-MADEED et al., 2014), como espaçamento, vales, angulações, dentre outras. A segunda abordagem busca representar a escrita através de descritores de textura, que extraem características globais, utilizando o documento manuscrito como um todo (BERTOLINI et al., 2013; HANNAD et al., 2016).

As bases de documentos manuscritos normalmente possuem um único estilo de escrita. Entretanto atualmente algumas bases utilizam mais de um estilo de escrita, ou seja, o mesmo escritor produz documentos em línguas diferentes, como, português, árabe, inglês, alemão, dentre outras (FREITAS et al., 2008; MÁADEED et al., 2012; KLEBER et al., 2013). Ao considerar mais de um estilo de escrita para o problema de identificação de escritores, a tarefa torna-se ainda mais árdua, pois é necessário construir um sistema de identificação independente do alfabeto utilizado.

A fim de solucionar o problema de identificação de escritores que produzem documentos em línguas diferentes, foram utilizados diversas técnicas como citado anteriormente. Uma delas é utilizar a rede neural convolucional (CNN) como extrator de características e classificador, visto que é uma característica da rede neural convolucional extrair as características por meio da convolução (NIELSEN, 2016). Além disso, com o intuito de aumentar as taxas alcançadas pela CNN e pelas características por ela extraídas, empregaremos a técnica de dissimilaridade. Sendo possível transformar qualquer problema de n-classes em um problema binário (HANUSIAK et al., 2011), e ajudando a contribuir ainda mais para o desempenho.

Assim, o foco deste trabalho é propor um método para o problema de identificação de

escritores usando documentos manuscritos *off-line*. Utilizaremos características extraídas pela rede neural convolucional, podendo ser consideradas como características globais. Avaliaremos o desempenho do método proposto em diferentes bases de dados, sendo elas texto-dependente e texto-independente. É válido ressaltar a utilização de duas abordagens que podem ser comparadas ao final deste trabalho, sendo elas, a utilização da técnica dissimilaridade e o desempenho dos classificadores utilizados.

## 1.1. Motivação

A principal motivação deste trabalho está na construção de um método semi-automático para solucionar o problema de identificação de escritores utilizando documentos manuscritos em mesma língua ou diferentes línguas, utilizando características extraídas pela rede neural convolucional.

Outro aspecto em relação a geração da textura é que a língua de cada documento se torna menos importante em relação a outras técnicas, uma vez que a técnica baseia-se somente nas características globais da escrita e não em características específicas da escrita, como por exemplo a existência ou não de laços e/ou vales.

Além disso, é válido ressaltar que a identificação é um problema extremamente desafiador, uma vez que existem diversas técnicas para melhorar o desempenho do problema, e mesmo assim é um campo de investigação em aberto, visto que nenhuma técnica consegue ser genérica suficiente para qualquer tipo de base de dados.

## 1.2. Desafios

Existem alguns desafios que podem ser encontrados na realização deste trabalho. O primeiro está relacionado ao tamanho da imagem como entrada para a rede neural convolucional, pois os exemplos que alcançam altas taxas de acerto, utilizam bases de dados com imagens pequenas. Exemplo: MNIST ( $28 \times 28$ ), STL-10 ( $96 \times 96$ ), Norb ( $108 \times 108$ ), Cifar-100 ( $32 \times 32$ ). (HAFEMANN, 2014).

Um outro aspecto é em relação a configuração da rede e dos parâmetros da rede neural convolucional. Não existe um método para ajustar os parâmetros, como também a configuração da rede, desta forma, é necessário basear-se na literatura e no método empírico de acordo com domínio da classificação.

Em relação as bases de dados utilizadas, temos o desafio de extrair características que sejam boas independente da língua escrita de determinada amostra, ou seja, criar um modelo robusto para todas as bases utilizadas.

### 1.3. Objetivos

O objetivo geral deste trabalho é avaliar a rede neural convolucional para extração automática de características e para classificação utilizando um sistema semiautomático de identificação *off-line* de escritores. Como objetivos específicos podemos destacar:

- Avaliar as características extraídas a partir da rede neural convolucional;
- Comparar o desempenho gerado utilizando a rede neural convolucional e utilizando as características extraídas pela CNN no *Support Vector Machine*;
- Avaliar o impacto da quantidade de blocos utilizadas no conjunto de treinamento e teste;
- Avaliar o impacto do tamanho dos blocos utilizados;
- Empregar a abordagem de dissimilaridade a fim de transformar um problema de *n-classes* em um problema binário;
- Comparar o desempenho das características com e sem o a abordagem de dissimilaridade;
- Construir um sistema semiautomático de identificação de escritores;
- Avaliar o desempenho do método proposto em diferentes bases de dados.

### 1.4. Estrutura do trabalho

O restante do trabalho está organizado da seguinte forma: O Capítulo 2 apresenta uma visão geral de alguns trabalhos sobre a identificação de escritores a partir de documentos manuscritos. No Capítulo 3 são apresentados alguns conceitos necessários para melhor entendimento do método empregado neste trabalho. O Capítulo 4 descreve o método que será utilizado para elaboração deste trabalho. O Capítulo 5 apresenta os experimentos realizados a partir do método utilizado. Por fim, o Capítulo 6 apresenta as conclusões levantadas durante o desenvolvimento deste trabalho.



---

## Estado da Arte

---

Neste capítulo apresentaremos uma revisão da literatura em relação ao processo de identificação de escritores em documentos manuscritos. Seleccionamos trabalhos que empregaram bases públicas conhecidas na literatura e também que utilizam técnicas computacionais similares as empregadas neste trabalho. Os trabalhos foram agrupados de acordo com a abordagem empregada, ou seja, por abordagens tradicionais de classificação e abordagens que utilizam aprendizagem profunda.

### 2.1. Abordagens Tradicionais de Classificação

A abordagem utilizada por Siddiqi e Vincent (2009) foi a extração de características por meio da técnica *Chain Code Histograms*, que se baseia no contorno do texto manuscrito. A base utilizada neste trabalho foi a IAM com 650 escritores, a qual utilizou apenas duas amostras por escritor, e dividiu ao meio as amostras de escritores que possuíam apenas uma amostra. Os experimentos utilizaram como métrica de classificação as distâncias Euclidiana,  $X^2$ , Bhattacharyya e de Hamming, além disso, a base de dados foi dividida em uma amostra para o conjunto de treinamento e outra para conjunto de teste. Os autores reportam taxa de identificação de 86% para o melhor caso, na qual empregou-se a distância  $X^2$ .

Bertolini et al. (2013) empregam em seu trabalho um método de geração de textura em documentos manuscritos. Utilizando as bases BFL e IAM em seus experimentos, as mesmas passam inicialmente por uma etapa de pré-processamento para a geração de textura e posteriormente essa textura é dividida em blocos de tamanhos fixos. A partir dos blocos de textura, foi empregado descritores (LBP e LPQ) de textura para extrair características de cada amostra. Por fim, foi utilizado uma abordagem de dissimilaridade com o intuito de converter um problema de n-classes para um problema binário. Utilizando o classificador SVM, foi avaliado o uso da dissimilaridade, o número de referências utilizado na dissimilaridade, o

número de classes utilizadas no conjunto de treinamento e o número de blocos. O número de escritores utilizados na base BFL foram 315, sendo 200 escritores para treino e 115 para teste, enquanto na base IAM foram utilizados 650 escritores, sendo 390 escritores para treino e 260 para teste. As melhores taxas de identificação apresentadas neste trabalho foram de 99,4% (BFL) e 99,6% (IAM) com o descritor de textura LPQ.

Djeddi et al. (2013) realizaram o reconhecimento de documentos manuscritos extraíndo características de *Grey Level Run Length* (GLRL). Neste trabalho, os autores utilizaram a base *Latin/Greek* a qual possui 126 escritores, sendo duas amostras escrita na língua inglesa e duas em grego. O uso de alfabetos distintos aumenta o nível de dificuldade neste trabalho. Os experimentos realizados utilizaram os classificadores SVM e KNN, com uma amostra para o conjunto de treinamento e outra para o conjunto de teste. Apresentando as taxas de identificação de 92,06% usando o mesmo estilo de escrita e 76,59% para estilos diferentes.

Al-Maadeed et al. (2014) propõem um sistema para identificar escritores utilizando características grafométricas. Foram utilizadas duas bases, a QUWI que possui 1017 escritores, sendo duas amostras em inglês e duas em árabe, e a IAM com 657 escritores, a qual utilizou apenas duas amostras por escritor. Na base IAM como alguns escritores cederam apenas uma amostra de documento manuscrito, divide-se ao meio a amostra para terem uma parte para treinamento e outra para teste. Todos os experimentos utilizaram o classificador *Kernel Discriminant Analysis* (KDA). Para base QUWI, o conjunto de treinamento possui dois parágrafos e no conjunto de teste possui um parágrafo da mesma amostra, e na base IAM utilizou uma amostra (documento) para treinamento e outra para teste. As taxas de identificação obtidas foram de 87,67% e 82,70%, para as bases QUWI e IAM, respectivamente.

Hannad et al. (2016) também utilizou uma abordagem de geração de textura na base de dados IAM empregando todos os 657 escritores. Neste trabalho, os autores utilizaram 60% das linhas dos documentos para o conjunto de treinamento e 40% para o conjunto de teste. As características foram extraídas pelos descritores de texturas LBP, LPQ e LTP. Os experimentos utilizaram como métrica de classificação a dissimilaridade calculada entre cada escritor, a partir da distância de Hamming. A melhor taxa de identificação reportada pelos autores foi de 89,54% utilizando o descritor LPQ.

No trabalho de He e Schomaker (2016), foi utilizado a base de dados CERUG, com 105 escritores que produziram quatro amostras cada, sendo duas em chinês, uma em inglês e uma em inglês e chinês. As características utilizadas foram LBPruns, na qual o *runs* é definido por uma sequência de pixels conectados com o mesmo valor na escala de cinza, LBPruns\_b, que é aplicado com a imagem binarizada, e COLD, que é uma características baseada no contorno e na curvatura da escrita. Os experimentos foram realizados utilizando a distância  $X^2$  como métrica de classificação, os melhores resultados foram 94,2% (LBPruns\_b) e 93,8% (COLD).

## 2.2. Abordagens utilizando Aprendizagem Profunda

Hafemann (2014), realizou experimentos utilizando o classificador CNN com as bases IAM e BFL. Foram avaliados diferentes parâmetros do CNN além do impacto do tamanho do filtro e o tamanho do bloco utilizado. O autor utilizou a representação de textura descrita por Bertolini et al. (2013) para ambas as bases de dados. Na base BFL foi utilizada 115 escritores e na base IAM foi utilizada 240. A partir da textura gerada, blocos de tamanhos  $256 \times 256$  foram criados. A dimensão do bloco avaliada pelos autores refere-se a utilizada pela CNN. As melhores taxas obtidas foram com o tamanho de filtro igual a  $7 \times 7$  e o tamanho do bloco igual a 48, sendo elas 91,67% (IAM) e 95,65% (BFL).

O trabalho de Fiel e Sablatnig (2015) utilizou uma abordagem diferente dos demais trabalhos apresentados anteriormente, a qual utilizou a base IAM para treinar a CNN e para o conjunto de teste foi utilizada a base CVL com 309 escritores. A partir do modelo da CNN foi extraído um vetor com 4096 características para cada amostra do conjunto de teste, e realizado os experimentos utilizando os vizinhos mais próximos como métrica de classificação. A taxa de identificação alcançada foi de 98,9%. Neste trabalho, os autores consideraram cinco amostras no conjunto de testes, sendo quatro amostras em inglês e uma em alemão.

No trabalho de Christlein et al. (2015), foi utilizado a base de dados CVL, com 310 escritores, que foi utilizado 27 escritores para treinar um modelo na qual cada escritor possuía sete amostras, e conjunto de teste com 283 escritores sendo cinco amostras cada. As características foram extraídas através da CNN, na qual o vetor de características foi submetido a técnica *Gaussian Mixture Model* (GMM) e posteriormente normalizado. Os experimentos foram realizados utilizando a métrica TOP- $k$  e mPA apresentando taxas de identificação de 99,4% para o melhor caso.

Em outro trabalho, Hafemann et al. (2016b) utilizou novamente a CNN, porém como extrator de características agora para o problema de verificação de assinaturas. A base utilizada foi a GPDS-960, que possui 881 escritores, com 24 assinaturas genuínas e 30 falsificações cada. As bases foram pré-processadas utilizando o algoritmo de Otsu, em seguida invertendo a intensidade dos pixels e por fim normalizando a entrada. Neste trabalho, foi utilizado uma abordagem de treinamento independente, a qual uma parte dos escritores e suas respectivas amostras são utilizadas apenas para o treinamento e outra parte para teste. As características foram extraídas pelo classificador CNN e classificadas pelo SVM, sendo o número total de características extraídas igual a 4096. Os experimentos utilizaram como conjunto de treinamento 581 escritores e para o conjunto de teste 300 escritores, a melhor taxa obtida foi de 97,83%.

## 2.3. Discussão

A Tabela 2.1 apresenta um resumo dos trabalhos detalhados anteriormente. É possível analisar que a grande maioria dos trabalhos possuem taxas de identificação acima 85%. Contudo, é difícil realizar uma comparação efetiva em relação ao desempenho dos trabalhos, já que na maioria dos casos utilizam bases diferentes ou então subconjuntos da base de dados. Outra característica comum é o uso de diferentes quantidades de texto manuscrito em cada trabalho, alguns autores utilizam somente palavras, outros algumas linhas, ou todo o documento, dentre as mais diferentes formas.

Tabela 2.1. Resumo do estado da arte.

Autores	Ano	Base de dados	Nível	Características	Classificação	Desempenho (%)
Siddiqi; Vincent	2009	IAM (650)	Documento Todo	<i>Chain code histograms</i>	Distância Euclidiana	86,00
Bertolini et al.	2013	IAM (650) BFL (315)	Blocos	Textura (LPQ)	SVM	96,70 99,20
Djeddi et al.	2013	Latin/Greek (126)	Documento Todo	GLRL	KNN SVM	73,41 73,02
Al-maadeed et al.	2014	IAM (657) QUWI (1017)	Parágrafo	Grafométricas	KDA	82,70 87,67
Hannad et al.	2016	IAM (657)	Linhas	Textura (LPQ)	Dissimilaridade (Distância de Hamming)	89,54
He e Schomaker	2016	CERUG (105)	Documento Todo	Textura (LBPruns_b)	Distância $X^2$	94,20
Hafemann	2014	IAM (240) BFL (115)	Blocos	CNN	CNN	91,67 95,65
Fiel; Sablatnig	2015	CVL (309)	Blocos	CNN	Distância $X^2$	98,90
Christflein et al.	2015	CVL (310)	Documento Todo	CNN e GMM	$mAP$ e $TOP-k$	99,40
Hafemann et al.	2016	GPDS-960 (881)	Documento Todo	CNN	SVM	97,83

---

## Fundamentação Teórica

---

Neste capítulo, será apresentado as principais técnicas computacionais a serem empregadas neste trabalho. A Seção 3.1 apresenta as bases de dados utilizadas neste trabalho. A Seção 3.2 realiza uma breve explicação da geração da textura. Na Seção 3.3 serão descritos conceitos de aprendizagem profunda e redes neurais convolucionais. Por fim a Seção 3.4 descreve a técnica de dissimilaridade.

### 3.1. Base de Dados

Três bases de dados poderão ser utilizadas neste trabalho, a primeira base é *Brazilian Forensic Letter* (BFL) (FREITAS et al., 2008), a segunda base é a QUWI (MÁADEED et al., 2012) e a terceira base é a CVL (KLEBER et al., 2013). Nas seções a seguir apresentaremos mais detalhes sobre estas bases de dados. Todas as três são bases públicas e encontram-se disponíveis para download.

#### 3.1.1. BFL

A base BFL é composta por 315 escritores, com três amostras por escritor em português do Brasil. As amostras foram fornecidas por alunos de graduação em três sessões diferentes ao longo de um mês. Os textos foram recolhidos em uma folha branca A4, e depois digitalizados em escala de cinza a 300dpi. Cada escritor utilizava sua própria caneta, portanto, existe uma grande variação nos documentos neste sentido. O conteúdo presente em cada documento é o mesmo, ou seja, texto-dependente. Neste trabalho utilizaremos 115 escritores para a abordagem da rede neural convolucional, escolhidos de forma aleatória, o mesmo subconjunto utilizado por Bertolini et al. (2013) e Hafemann (2014). A Figura 3.1 apresenta um documento manuscrito da base BFL.

### 3.1.2. QUWI

A base QUWI possui 4068 documentos manuscritos digitalizados a 600dpi. A base conta com 1017 escritores na qual cada escritor cede quatro amostras. A base é dividida em dois idiomas, os dois primeiros documentos são escritos em árabe e os demais em inglês. Os documentos possuem texto-dependente e texto independente, o conteúdo do primeiro e do terceiro documento são textos-independente (muda de escritor para escritor), e o segundo e quarto documento possuem o mesmo conteúdo. Neste trabalho utilizaremos um subconjunto da base contendo 475 escritores, que foi utilizada por Hassaine et al. (2013). A Figura 3.2 apresenta um documento manuscrito da base QUWI.

De  
 Fernando Quintas Zanon  
 Rua Luiz Kyré da Kerez, 87 - Ap. 300  
 Xenópolis, Nova York 14506-158

Para  
 Dr. Onório Bob Grant

Soube, através de publicação pela imprensa local, que V. Ss. necessitam de um funcionário na Seção de Correspondência do Departamento FENAL. Venho, portanto, candidatar-me a esta vaga sou brasileiro, solteiro, com 18 anos, curso a 3ª série do Curso Técnico de Contabilidade do Colégio Honório Alves - Escola Municipal de 2º Grau - e possui alguma prática de ditilografia e Arquivos.

Trabalhei durante dois anos nas Lojas Universais Rayon SA onde exerci as funções de Auxiliar de Escritório Júnior. Inicialmente, coloco-me à disposição de V. Ss. para um período de experiência requerido, entao, poderei tranquilamente avaliar minhas aptidões.

Na expectativa de uma resposta apresento-lhes cordiais Saudações

Fernando Zanon

Figura 3.1. Documento da base BFL.

كانت عناصر اقتصادية او سياسية او اجتماعية او ذات طابعه دينيه ،  
 ولا توجد اي قاعدة بيانات او ارشيف للبرامج الجرافيه عن النساء في الثقافات  
 الارابيه لكن هناك المتخصصات الجرافيه المسماة " بيوجرافيك بيرسبيكتيف او  
 وعين ، اي (رؤى جرافيه عن السلوك) التابعه لجمعية الجغرافيين الامريكانيين  
 تكفل بقاعدة بيانات موجهة على شبكة الانترنت في الموقع الالكتروني في  
 شبكة المعلومات كما توجد مقالات عن السلوك في الثقافات الارابيه

**Figura 3.2.** Documento em árabe com escrita livre da base QUWI.

### 3.1.3. CVL

A base CVL é composta por 309 escritores, sendo 1609 documentos manuscritos com texto-dependente, digitalizados a 300 dpi. Além disso, 27 escritores possuem 6 documentos em inglês e 1 em alemão, enquanto, 283 escritores possuem 4 documentos em inglês e 1 em alemão. O fato de a base ser bi-script é interessante pois podemos avaliar o impacto de possuir mais de um estilo de escrita no conjunto de teste. A Figura 3.3 apresenta um documento manuscrito da base CVL.

And fortune, on his damnd quorrel smiling,  
 Show'd like a rebel's whore: but all's too weak:  
 For Iwan Nomace - well he deserves that name -  
 Disdaining fortune, with his brandish'd steel,  
 Which smok'd with bloody execution,  
 Like valour's minion carv'd out his passage  
 Till he faced the slave;

**Figura 3.3.** Documento em inglês da base CVL.

## 3.2. Geração de Textura

O processo de geração de textura proposto por Hanusiak et al. (2011) é uma técnica que consiste em eliminar os espaços em brancos entre as palavras e as linhas, a partir da imagem original. Esta técnica é utilizada para conseguir gerar uma textura mais densa, de forma que represente melhor uma escrita.

A técnica consiste em binarizar a imagem utilizando o algoritmo de Otsu (OTSU, 1979), e depois percorrer a imagem para localizar os componentes conexos, as partes que se interligam ao decorrer da escrita, na qual componentes conexos pequenos demais são desconsiderados evitando os possíveis ruídos. As caixas delimitadoras dos demais componentes

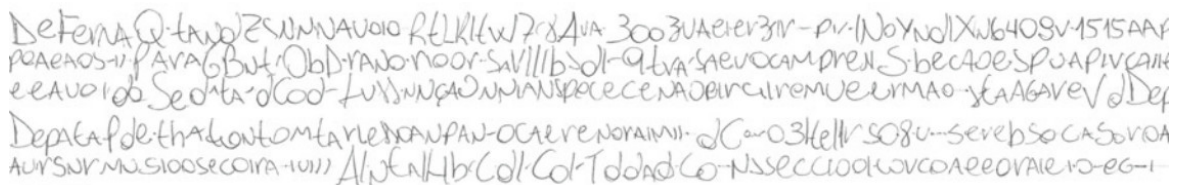


conexos são utilizadas para extrair os componentes das imagens originais em escala de cinza. Por fim, os componentes conexos são alinhados gerando uma nova imagem utilizando o centro de massa da caixa delimitadora, a Figura 3.4a exemplifica este processo. Depois de preencher a primeira linha, é calculada a altura média de todos os componentes utilizados. A Equação 3.1 é utilizada para definir a coordenada  $y$  da linha seguinte. Na Equação 3.1 a divisão de  $h$ , que é altura média das caixas delimitadoras, contribui para diminuição da distância entre as linhas e constrói texturas mais representativas, caso contrário, a textura ainda apresentaria grande conteúdo de espaços em brancos, como na Figura 3.4b (HANUSIAK et al., 2011).

$$y_{new} = y_{previous} + \frac{h}{2} \quad (3.1)$$



(a)



(b)

**Figura 3.4.** Geração de textura. (a) Alinhamento dos componentes e (b) a textura. Retirada de Hanusiak et al. (2011).

A abordagem proposta por Hanusiak et al. (2011) torna-se interessante pois geramos uma textura mais densa e representativa do escritor. Mesmo removendo informações que podem contribuir para identificação de um escritor, como por exemplo o espaçamento que o mesmo usa em seu texto. Este processo de geração de textura apresentou desempenho superior ao comparar com o uso do documento original (BERTOLINI et al., 2013).

### 3.3. Aprendizagem Profunda

A aprendizagem profunda é uma subárea de aprendizagem de máquina que está sendo muito utilizada em diversas aplicações a partir de filtragem de conteúdo em redes sociais para realizar recomendações em sites de *e-commerce* (LECUN et al., 2015). A principal característica da aprendizagem profunda é a utilização de camadas para realizar a extração de características e classificação (LECUN et al., 2015).

Segundo LeCun et al. (2015), as principais técnicas de classificação precisavam de um pré-processamento em relação aos dados de entrada como também a extração de características. Estas duas etapas podem influenciar fortemente o desempenho do sistema, pois se uma delas falhar, o sistema de classificação pode ser comprometido.

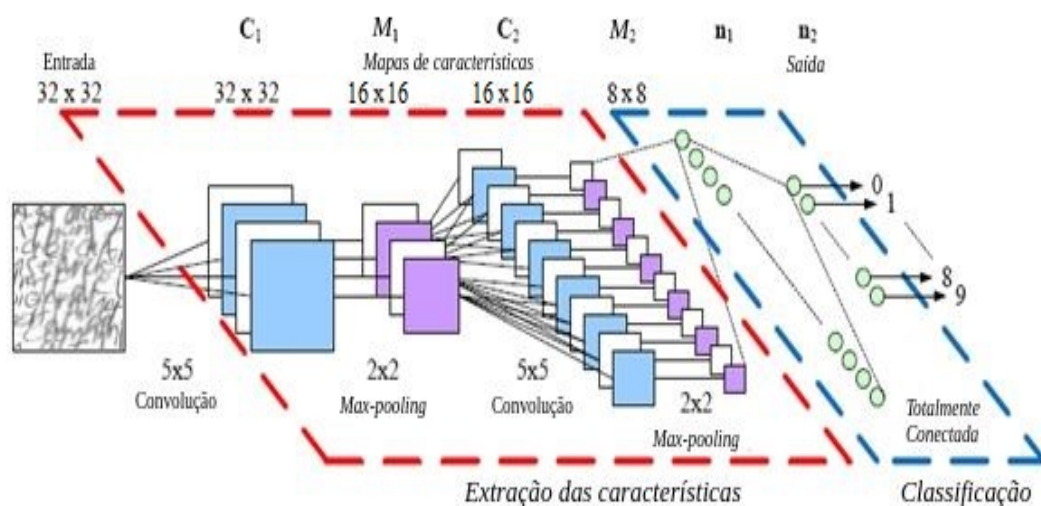
Um dos fatores que torna a aprendizagem profunda ainda mais interessante, é que

a mesma descarta estas etapas, ou seja, é possível conseguir um desempenho similar aos apresentados em literatura sem a necessidade de qualquer tipo de pré-processamento ou estudos aprofundados em métodos de extração de características.

### 3.3.1. Rede Neural Convolutiva (CNN)

Redes neurais convolucionais são biologicamente inspiradas e capazes de serem treinadas e aprenderem representações invariantes a escala, translação, rotação e transformações (JARRETT et al., 2009). A rede neural convolutiva utiliza um sistema hierárquico que tenta representar a estrutura em relação ao reconhecimento de uma imagem, na qual pixels formam arestas, arestas formam padrões, que formam objetos, que por sua vez descrevem as cenas (AREL et al., 2010).

Segundo Abdel-Hamid et al. (2012) a CNN consiste em um ou mais pares de camadas de convolução e *max-pooling*. As camadas convolucionais aplicam filtros que processam pequenos locais de uma imagem, e são replicados por toda imagem. As camadas de *max-pooling* geram uma versão com menor resolução das camadas de convolução aplicando a ativação máxima do filtro em diversas posições dentro de uma janela. Assim, é adicionado mais tolerância para regiões específicas de um determinado objeto na imagem. As camadas mais altas utilizam filtros que funcionam a partir de entradas de baixa resolução para processar as partes com maior complexidade da imagem. Por fim, a camada totalmente conectada combina as entradas de todas as posições para realizar a classificação das entradas globais (ABDEL-HAMID et al., 2012).



**Figura 3.5.** Exemplo de arquitetura da CNN. Adaptada de Ward et al. (2011).

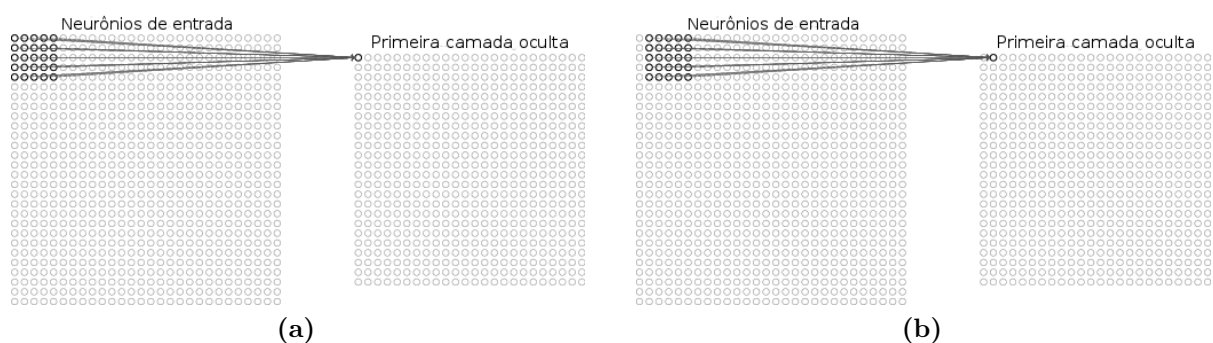
A Figura 3.5 apresenta um exemplo de arquitetura da CNN, tendo como entrada uma imagem de tamanho  $32 \times 32$  e posteriormente uma camada convolutiva com tamanho

do campo receptivo local (filtro) igual a  $5 \times 5$ , e uma camada de *max-pooling* com tamanho do filtro de  $2 \times 2$ . Em seguida executa as duas camadas citadas anteriormente com a mesma configuração, por fim existem duas camadas totalmente conectadas, a qual a última possui o número de neurônios igual a quantidade de classes existentes no problema. Desta forma, o exemplo apresenta um problema com dez classes e na camada convolucional não há perdas da imagem, pois neste caso é utilizado uma abordagem de preencher as perdas (bordas) com zero, na qual também utilizaremos neste trabalho (KARN, 2016). A seguir, apresentaremos com maior detalhes as camadas convolucionais, *pooling* e totalmente conectada.

### 3.3.2. Camada Convolucional

A camada convolucional possui filtros que são aplicados por toda entrada, na qual, para cada filtro, um neurônio está ligado apenas a um subconjunto dos neurônios na camada anterior. Os filtros geram mapas de características a partir de um pequena região chamada campo receptivo local. Além disso, existe pesos compartilhados entre os neurônios, levando o filtro aprender padrões frequentes em qualquer parte da entrada (HAFEMANN, 2014).

- **Campos Receptivos Locais:** Na camada convolucional, cada neurônio corresponde a intensidade de cada pixel da imagem, na qual uma região que representa o tamanho do filtro utilizado que é conectado a um único neurônio da camada oculta. Esta região é denominada de campo receptivo local, a qual cada conexão aprende um peso e cada neurônio aprende um viés global (NIELSEN, 2016), ilustrada na Figura 3.6a. Em seguida, esta região é percorrida pela imagem inteira, na qual, cada deslocamento é representado por um novo neurônio da camada oculta, visto na Figura 3.6b. A Figura 3.6 apresenta uma imagem de tamanho  $28 \times 28$  e o campo receptivo local de tamanho  $5 \times 5$ .



**Figura 3.6.** Exemplo do campo receptivo local. Adaptada de Nielsen (2016).

- **Pesos Compartilhados:** Os pesos compartilhados dentro da camada convolucional são os mesmos para cada campo receptivo local, isto quer dizer que os neurônios da primeira camada oculta detectam o mesmo padrão, porém em outras regiões da

imagem. Esta é uma característica que torna a CNN adaptativa em relação a diferentes representações que um padrão possa ter (NIELSEN, 2016).

- **Filtros:** Os filtros citados anteriormente são os responsáveis pela formação dos mapas das características, ou seja, na camada convolucional se aplica o filtro realizando uma convolução por toda imagem. Realizando operações como detecção de borda, nitidez e borrão apenas mudando os valores numéricos da matriz de filtro antes da operação de convolução, isso significa que diferentes filtros podem detectar diferentes recursos de uma imagem, por exemplo bordas, curvas, etc (KARN, 2016). A Figura 3.7 apresenta exemplos de alguns filtros que podem ser utilizados.

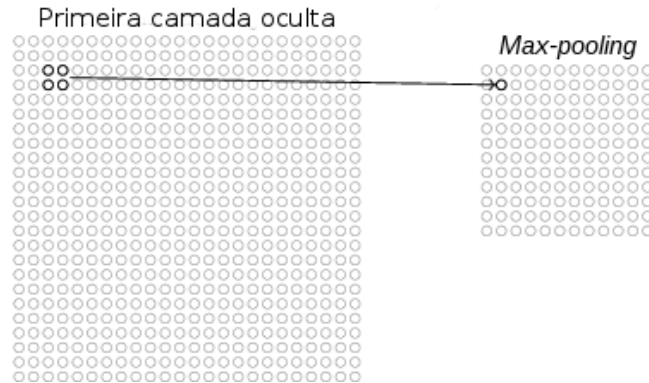
Operação	Filtro	Imagem Convolucionada
Identificação	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Detecção de Borda	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Nitidez	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Desfoque	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Desfoque Gaussiano	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Figura 3.7. Exemplos de filtros da camada convolucional. Adaptada de Karn (2016).

### 3.3.3. Camada de Pooling

Em geral, a camada de *pooling* é utilizada posteriormente a camada convolucional e sua função é diminuir o tamanho da imagem para encontrar possíveis padrões que possam ficar

mais evidentes (NIELSEN, 2016). Um procedimento comum na camada de *pooling* é chamado de *max-pooling*, na qual em uma unidade de *pooling* exibe-se a ativação máxima na região de entrada (NIELSEN, 2016). Na Figura 3.8 podemos ver um exemplo de *max-pooling* em uma região  $2 \times 2$  em relação a região da camada convolucional, portanto a imagem é redimensionada para  $12 \times 12$ .



**Figura 3.8.** Exemplo de *max-pooling*. Adaptada de Nielsen (2016).

### 3.3.4. Camada Totalmente Conectada

A camada totalmente conectada é a camada que também está presente nas redes neurais comuns, ela é responsável por conectar as camadas sem utilizar pesos compartilhados (HAFEMANN, 2014).

Além disso, o termo totalmente conectada implica que cada neurônio na camada anterior está conectado a cada neurônio na próxima camada e ela adiciona uma camada de saída com o número de neurônios em relação ao número de classes presente no respectivo experimento, para realizar a classificação utilizando a função de ativação softmax (KARN, 2016).

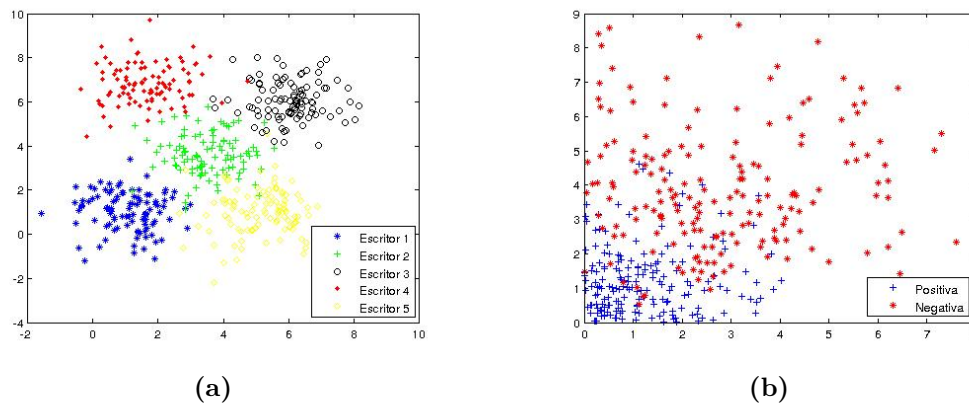
Além da classificação, adicionar uma camada totalmente conectada também é uma forma barata de aprender combinações não-lineares desses recursos. A maioria dos recursos das camadas convolucionais e de pool pode ser bom para a tarefa de classificação, mas combinações dessas características podem ser ainda melhores (KARN, 2016).

## 3.4. Dissimilaridade

A dissimilaridade segundo P'kalska e Duin (2000) é uma técnica que transforma um problema de n-classes em um problema binário. Portanto, a técnica produz um sistema robusto quando existe um pequeno número de amostras para um número grande de classes (MARTINS et al., 2015).

Neste trabalho iremos utilizar a mesma abordagem apresentada por Hanusiak et al. (2011), cuja a ideia geral é realizar o cálculo da dissimilaridade a partir do vetor de características. Portanto, o problema multi-classe será transformado em um problema binário, e conseqüentemente será criada duas classes que serão independentes do número de escritores, sendo elas, positiva, que representa uma única classe (autor) e a negativa, que representa as demais classes (não autor).

A Figura 3.9 ilustra um exemplo da transformação citada anteriormente, sendo a Figura 3.9a a representação do espaço de características de várias instâncias de cinco diferentes escritores. Na Figura 3.9b temos a representação do espaço de dissimilaridade, das instâncias dos cinco escritores, assim a classe positiva representa as amostras genuínas de cada escritor e a classe negativa representa as amostras negativas. Portanto, a classe positiva é representada pela diferença entre instâncias da mesma classe, enquanto para a classe negativa, é calculada a diferença de instâncias de uma classe com outras diferentes classes. Assim, deseja-se para a classe positiva um vetor com valores próximo de zero, enquanto para as classes negativas um vetor com valores distante de zero. É válido ressaltar que a base de dados utilizada pela Figura 3.9 é sintética, e meramente ilustrativa, ou seja, cada escritor é representado por um valor  $x$  e  $y$ .

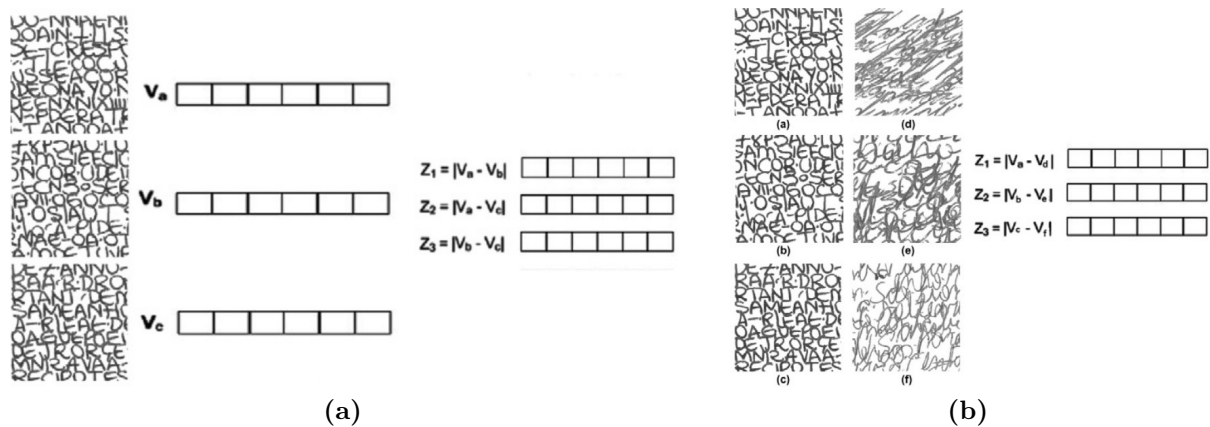


**Figura 3.9.** Exemplo de dissimilaridade utilizando uma base de dados sintética.

O cálculo do vetor de dissimilaridade é realizado aos pares e realiza-se a diferença entre os vetores de características de uma instância de classe com instâncias da mesma classe (positiva) e classes diferentes (negativa). Portanto, para a classe positiva a diferença é realizada apenas entre as amostras da mesma classe, enquanto para a classe negativa é realizado entre amostras da classe de interesse e amostras das outras classes. O número de referências representa a quantidade de amostras aleatórias que será utilizado para combinação em pares. A Equação 3.2 representa o cálculo realizado para a combinação, com  $n$  elementos distintos, agrupados  $p$  a  $p$ . Assim, com três referências, temos 3 combinações positivas e 3 negativas, com 5 temos 10 e 10, com 9 temos 36 amostras positivas e 36 negativas.

$$C_{n,p} = \frac{n!}{p!(n-p)!} \quad (3.2)$$

A Figura 3.10 ilustra o processo de dissimilaridade a partir das características extraídas de cada bloco da textura. A Figura 3.10a mostra o processo para geração da classe positiva, a qual são gerados três vetores de dissimilaridade a partir dos vetores de características extraídas dos blocos da mesma classe, e na Figura 3.10b é realizado o processo para a classe negativa, que utiliza os vetores de características dos blocos da classe positiva, porém a diferença é efetuada com os vetores de características de blocos de outras classes, escolhidas aleatoriamente.



**Figura 3.10.** Dissimilaridade (a) entre amostras do mesmo escritor para gerar amostras positivas e (b) entre amostras do escritor da classe positiva e diferentes escritores para gerar amostras negativas. Adaptado de Hanusiak et al. (2011).

### 3.5. Combinação de Classificadores

A combinação de classificadores é apresentada por (KITTLER et al., 1998), na qual o seu objetivo é a combinação das predições dada pelos classificadores. Neste trabalho a proposta é identificar o escritor de um determinado documento manuscrito. Como o documento foi transformado em uma textura e posteriormente fragmentado em diversos blocos, utilizaremos a combinação para termos uma decisão final a partir da predição dos  $n$  blocos de um determinado escritor. Assim, todos os documentos serão divididos em blocos de tamanho  $m \times m$  a qual o tamanho do bloco impacta diretamente no número dos mesmos.

Segundo os testes realizados por Bertolini et al. (2013) a regra descrita por Kittler (KITTLER et al., 1998) com melhor desempenho e o que empregaremos neste trabalho é a regra da soma, representada pela Equação 3.3, a qual  $Z$  é a predição final,  $\omega_j$  é a classe,  $x_i$  é um determinado bloco e  $P(\omega_k|x_i)$  é a probabilidade referente a  $x_i$  pertencer a classe  $\omega_j$ .

$$\text{Atribui } Z \rightarrow \omega_j \quad \text{se,} \quad \sum_{i=0}^n P(\omega_j|x_i) = \max_{k=1}^m \sum_{i=0}^n P(\omega_k|x_i) \quad (3.3)$$

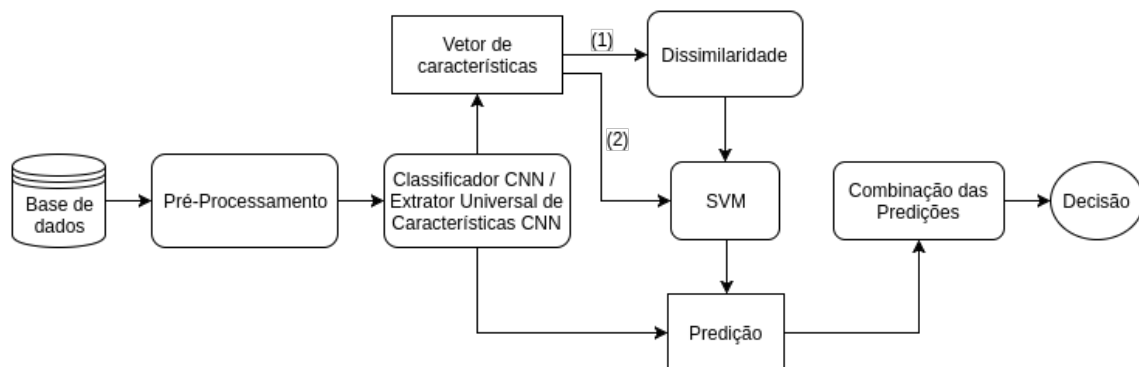
### **3.6. Considerações**

Neste capítulo realizamos uma breve explicação de algumas técnicas que serão utilizadas neste trabalho. A ideia é dar ao leitor uma introdução ao tema abordado. Para mais informações, consultar as referências bibliográficas. No próximo capítulo apresentaremos o método proposto neste trabalho.



## Método

Neste capítulo apresentaremos o método utilizado para a elaboração deste trabalho. A Figura 4.1 ilustra cada etapa a ser realizada para o desenvolvimento da abordagem proposta.



**Figura 4.1.** Visão geral do método proposto.

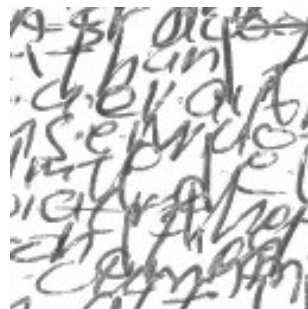
A partir da base de dados, foi realizado um pré-processamento nos documentos manuscritos de cada escritor com o objetivo de diminuir os ruídos existentes, como também gerar uma textura mais representativa de cada amostra. Posteriormente a imagem gerada a partir do documento original foi dividida em blocos de tamanho  $m \times m$ . Na etapa seguinte com uso do classificador CNN / extrator universal de características utilizaremos dois recursos providos pela CNN, sendo a classificação de cada bloco de entrada, ou seja, o processo tradicional do classificador CNN e também o uso da CNN como um extrator de características. Em seguida, foi realizado o cálculo da dissimilaridade sobre os vetores de características de cada amostra (1). Entretanto, poderemos avaliar o desempenho das características extraídas em outros classificadores, como o SVM (2). Na etapa seguinte foi utilizado o classificador SVM para classificar os blocos de textura, o qual apresentará uma probabilidade de confiança do bloco pertencer a determinado escritor. Por fim, foi realizado a combinação das predições geradas de cada bloco, pelo SVM e pela CNN, a fim de obter

uma decisão final sobre quem escreveu determinado documento. A seguir, todas as etapas serão descritas em suas respectivas seções.

## 4.1. Pré-Processamento

O pré-processamento utilizado neste trabalho pode ser dividido em geração de textura e divisão em blocos. Primeiramente é realizada a geração de textura de cada documento, esta abordagem é utilizada para conseguir representar melhor a escrita, tornando-a mais densa e representativa, e é descrita em detalhes na Seção 3.2. Esta abordagem produz bons resultados utilizando descritores de textura (BERTOLINI et al., 2013; HANUSIAK et al., 2011), e também ao utilizar o classificador CNN (HAFEMANN, 2014).

Em seguida, é realizado um esquema de divisão em blocos de cada imagem de textura. Os blocos são gerados de forma aleatória com sobreposição de conteúdo. A divisão é realizada de acordo com o objetivo do experimento, sendo que o impacto do número de blocos e o tamanho dos mesmos será investigado neste trabalho. Esta abordagem é utilizada para transformar uma amostra de textura em  $n$  blocos de tamanho  $m \times m$  que representam esta mesma amostra. Além disso, esta representação pode evitar o *overfitting* presente no treinamento da CNN, que será melhor detalhado na Seção 4.2. A Figura 4.2 apresenta um exemplo de bloco gerado após a geração de textura.



**Figura 4.2.** Bloco de tamanho  $128 \times 128$ .

## 4.2. Rede Neural Convolutacional

Neste trabalho são utilizados dois recursos presentes na rede neural convolutacional, sendo o classificador e o extrator universal de características, porém, é válido ressaltar que os dois recursos possuem o mesmo processo de treinamento e classificação.

Um dos grandes problemas ao treinarmos uma rede é o *overfitting*, na qual o modelo se torna específico demais para uma determinada base de dados. Uma forma de evitar este problema é realizar a divisão em blocos como citado anteriormente e a utilização de uma técnica chamada *dropout* (HINTON et al., 2012), que consiste em desativar um neurônio oculto a partir de uma probabilidade (KRIZHEVSKY et al., 2012). Segundo Krizhevsky et

al. (2012) esta técnica consegue deixar neurônios menos sobre ajustados, o que implica em forçar a rede aprender características mais robustas. Empregaremos ambas as técnicas neste trabalho.

### 4.2.1. Classificador

A partir da base de dados pré-processada utilizaremos um documento (ou parte dele) para o treinamento e outro para o teste. Em seguida será necessário configurar a arquitetura da rede, ou seja, configurar quais camadas serão inseridas, (Seção 3.3.1) e seus respectivos parâmetros:

- **Kernel Size:** É a quantidade de filtros que uma determinada camada possui ;
- **Filter Size:** É o tamanho do filtro que será aplicado sobre a entrada ;
- **Stride:** Distância entre as aplicações dos filtros.

Por fim, será necessário configurar os parâmetros da CNN, tendo como parâmetro principal o **Batch Size**, que representa a quantidade de amostras que será processada por vez na fase de treinamento. Assim podemos realizar o processo de classificação, na qual ao final do processo será gerado um modelo de predição.

Para avaliação do modelo utilizaremos a mesma abordagem apresentada por Hafemann (2014), que utilizou a *Accuracy*, visto na Equação 4.1.

$$Accuracy = \frac{Predições\ Corretas}{Amostras\ no\ conjunto\ de\ teste} \quad (4.1)$$

### 4.2.2. Extrator Universal de Características

Nesta etapa, utilizaremos o modelo gerado a partir do processo de treinamento a fim de obter as características extraídas pela CNN representadas na penúltima camada. As características são extraídas pelos filtros presentes em cada camada e o número de características estão relacionadas ao número de neurônios presentes na camada totalmente conectada, considerando a técnica *dropout*. Usualmente os trabalhos encontrados na literatura utilizam 4096 características (FIEL; SABLATNIG, 2015; HAFEMANN et al., 2016b), entretanto o impacto do número de características extraídas na taxa de identificação do sistema, será investigado neste trabalho.

## 4.3. Dissimilaridade

A partir do vetor de características extraído pela CNN na Seção 4.2.2 será realizado o cálculo da dissimilaridade. Esta técnica baseia-se no número de referências (blocos disponíveis) em relação ao cálculo realizado para a geração da classe positiva e negativa, descrito na Seção 3.4. Neste trabalho utilizaremos os mesmos números de referências que Bertolini et al. (2013). A

fim de avaliar o impacto do número de referências na classificação, a ideia é variar o número de referências, e analisar a robustez do sistema usando mais e menos referências.

#### 4.4. Máquina de Vetores de Suporte (SVM)

O classificador SVM será utilizado para realizar a classificação a partir das características extraídas através da CNN. No trabalho proposto investigaremos a robustez das características e também a abordagem de dissimilaridade, portanto, no primeiro caso temos um problema multi-classe na qual as classes representam o  $\omega$  escritores presentes nas bases de dados. Utilizaremos um modelo um-contra-todos do SVM com *kernel* Gaussiano. Os parâmetros  $C$  e  $\gamma$  serão otimizados através de busca exaustiva na qual utilizaremos o conjunto de treinamento para treinar parâmetros do SVM e um conjunto de validação é definido para avaliarmos o desempenho do sistema.

A utilização do classificador SVM deve-se ao fato do mesmo ser utilizado em diversos trabalhos (BERTOLINI et al., 2013; DJEDDI et al., 2013; HAFEMANN et al., 2016b), além do mesmo apresentar bons resultados para problemas binários, ou seja, tornando-se ideal para a dissimilaridade. Segundo Lorena e Carvalho (2007) as principais características do classificador SVM são a capacidade de generalização e de não ficar preso em um mínimo local, diferente de outras técnicas.

#### 4.5. Combinação das predições

A combinação das predições é realizada a fim de obter uma predição em relação a um documento todo, portanto, combina-se as predições de acordo com o número de blocos que um documento foi dividido, na qual é melhor detalhada na Seção 3.5. Esta combinação será realizada para todas as abordagens presentes neste trabalho. Visto que a textura é fragmentada em vários pequenos blocos.

#### 4.6. Considerações

Neste capítulo realizamos uma explicação detalhada de cada etapa do método proposto neste trabalho. No próximo capítulo apresentaremos os experimentos preliminares realizados.

---

## Experimentos

---

Neste capítulo descrevemos os experimentos e os resultados alcançados. Em nossos experimentos empregamos três diferentes bases de dados: BFL, CVL e QUWI descritas no Capítulo 3.

Inicialmente os experimentos utilizaram 115 escritores escolhidos de forma aleatória, pois o sistema possuía uma limitação de *hardware* em relação a CNN, e a divisão entre o conjunto de treinamento e de teste foi realizada de forma específica para cada base.

Todos os experimentos foram realizados empregando o *framework Keras* (CHOLLET, 2015) para a configuração e execução da CNN. Os principais parâmetros utilizados foram o número de *epoch* igual a 40, o *batch size* igual a 48 e a configuração da arquitetura demonstrada na Tabela 5.1, na qual  $\omega$  representa o número de classes utilizadas e  $n$  representa o tamanho da penúltima camada. Enquanto, para a execução da classificação com o SVM (CHANG; LIN, 2011) foi utilizado um esquema de um-contra-todos com *kernel Radial Basis Function* (RBF).

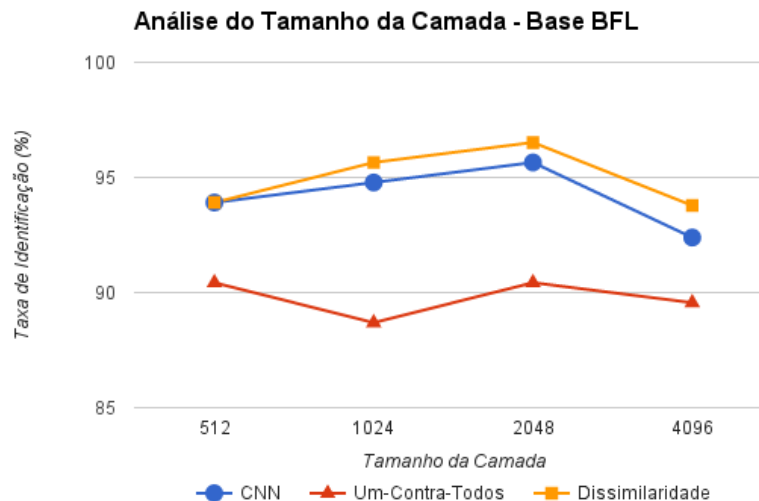
Os três primeiros experimentos realizados foram com o intuito de analisar o impacto que a quantidade de características, representada pela penúltima camada da CNN, o tamanho dos blocos e a quantidade de blocos podem contribuir para o desempenho do sistema. A ordem da realização nestes experimentos foram realizadas de acordo com o impacto no resultado obtido em experimentos preliminares.

Desta forma, os demais experimentos devem empregar os parâmetros que descrevem melhores desempenhos, sendo que cada experimento possui três abordagens: CNN, que utiliza suas características e seu próprio classificador, Um-Contra-Todos, que utiliza as características extraídas pela CNN e as classifica usando o SVM, e por fim, a Dissimilaridade que também utiliza as mesmas características porém é aplicado a abordagem de dissimilaridade que também é classificada usando o SVM. Além disso, as abordagens Um-Contra-Todos e Dissimilaridade utilizam somente nove blocos aleatórios dentre os blocos utilizados pela CNN.

**Tabela 5.1.** Configuração da arquitetura utilizada pela CNN.

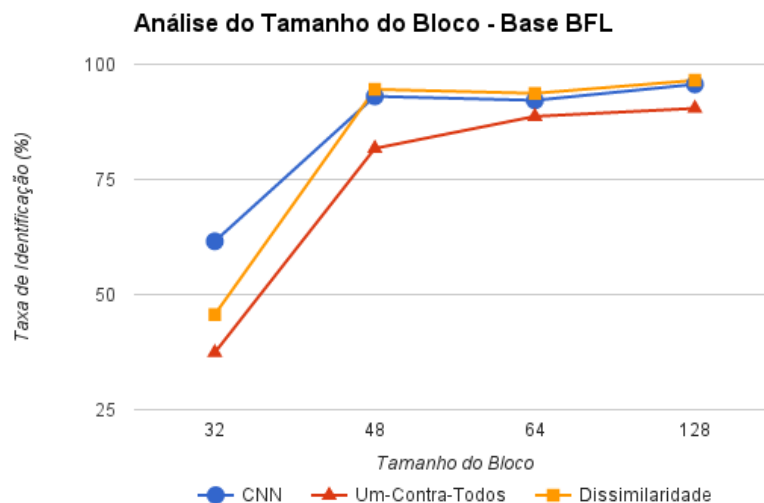
Camada	Quantidade de Filtros	Tamanho do Filtro	Outros Parâmetros
Convolutacional 1	32	3x3	<i>stride</i> = 1
Max-Pooling 1	-	2x2	<i>stride</i> = 2
Convolutacional 2	32	3x3	<i>stride</i> = 1
Convolutacional 3	64	3x3	<i>stride</i> = 1
Max-Pooling 2	-	2x2	<i>stride</i> = 2
Convolutacional 4	64	3x3	<i>stride</i> = 1
Max-Pooling 3	-	2x2	<i>stride</i> = 2
Convolutacional 5	128	3x3	<i>stride</i> = 1
Convolutacional 6	128	3x3	<i>stride</i> = 1
Max-Pooling 4	-	2x2	<i>stride</i> = 2
Totalmente Conectada + Dropout	$n$	-	$p = 0,5$
Totalmente Conectada	$\omega$	-	-

O primeiro experimento foi realizado a fim de avaliar o número de neurônios utilizado na penúltima camada pela CNN, que representa o número de características utilizados posteriormente pelo SVM, sendo  $n = \{512, 1024, 2048, 4096\}$ . Para este experimento utilizamos a base BFL com 115 escritores, na qual cada amostra (textura) foi dividida em 100 blocos de tamanho  $128 \times 128$  para a CNN, enquanto para as abordagens Um-Contra-Todos e Dissimilaridade foi utilizado nove blocos como citado anteriormente. Para o conjunto de treinamento foi utilizado as duas amostras (1,2), e no conjunto de teste foi utilizado a terceira. Portanto, no conjunto de treinamento temos 200 blocos referentes as duas primeiras amostras e 100 blocos no conjunto de teste referente a terceira amostra. A Figura 5.1 apresenta as taxas de identificação obtidas em cada abordagem citada anteriormente.

**Figura 5.1.** Experimentos realizados com a base BFL para analisar o desempenho em relação ao tamanho da penúltima camada.

Analisando os resultados apresentados através da Figura 5.1, podemos notar que a utilização de 512 e 4096 neurônios obtiveram resultados um pouco abaixo em relação aos demais. O fato de 512 neurônios obter pior taxa de identificação deve-se a limitação de características que pode não ser suficiente para distinguir as classes em comparação com as demais, e em relação a 4096 neurônios pode existir o problema de *overfitting*, que é quando as características ficam específicas demais para as classes do treinamento. Por fim, a utilização de 1024 e 2048 neurônios obtiveram os melhores resultados, sendo que 2048 apresentou resultados um pouco melhor. Tais números são similares as apresentadas por Hafemann et al. (2016a) (aceito para publicação).

No segundo experimento empregamos a melhor configuração no experimento anterior, portanto, todos os experimentos seguintes utilizarão 2048 neurônios na penúltima camada. Assim, neste experimento também utilizamos a base BFL, com 115 escritores, na qual cada amostra foi dividida em 100 blocos de tamanho  $m \times m$ , sendo  $m = \{32, 48, 64, 128\}$ . Empregamos os mesmo valores utilizados anteriormente para treinamento e teste. A Figura 5.2 apresenta a taxa de identificação para cada abordagem.

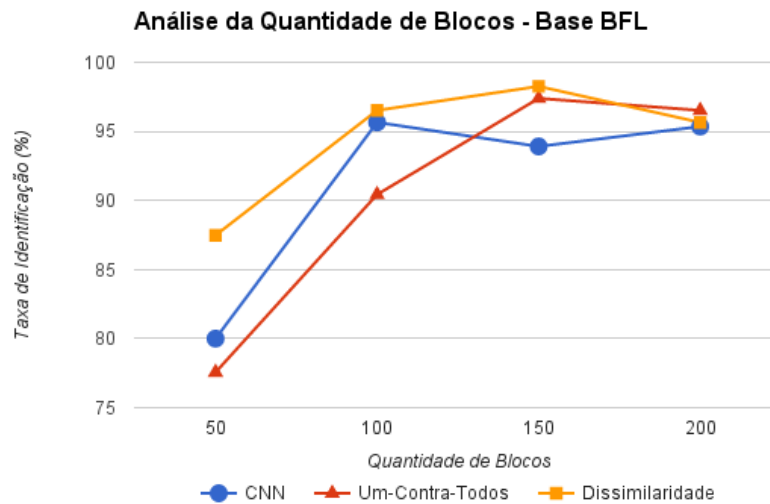


**Figura 5.2.** Experimentos realizados com a base BFL para analisar o desempenho em relação ao tamanho do bloco.

A partir dos resultados apresentados através da Figura 5.2, nota-se que quando é utilizado blocos com tamanho  $32 \times 32$  a taxa de identificação é bem menor em relação aos outros tamanhos, provavelmente isto ocorre devido ao tamanho do filtro ( $3 \times 3$ ) e a quantidade de camadas de *max-pooling* de tamanho  $2 \times 2$  utilizado pela arquitetura da rede. A cada *max-pooling* realizado as dimensões da imagem são divididas por dois, conseqüentemente dificultando a tarefa do extrator de características encontrar padrões nos blocos de cada amostra. Além disso, é possível analisar que quanto maior o tamanho do bloco melhor a taxa de identificação, e mais próximos são os resultados das três abordagens, na qual todos os

casos a abordagem de dissimilaridade apresentou melhor desempenho, tendo como  $128 \times 128$  o melhor desempenho.

No terceiro experimento também será utilizado as melhores configurações dos últimos dois experimentos, sendo eles 2048 neurônios na penúltima camada e o tamanho do bloco igual a  $128 \times 128$ . Na qual cada amostra foi dividida em  $k$  blocos de tamanho  $128 \times 128$ , sendo  $k = \{50, 100, 150, 200\}$ . A Figura 5.3 apresenta a taxa de identificação para cada abordagem.



**Figura 5.3.** Experimentos realizados com a base BFL para analisar o desempenho em relação a quantidade de blocos.

Ao realizarmos uma análise sobre o resultado da Figura 5.3, percebemos que quando cada amostra é dividida em 50 blocos a taxa de identificação têm pior desempenho em relação as outras configurações, uma possível explicação se encontra no treinamento da CNN, que necessita de uma grande quantidade de amostras para se obter melhores resultados (KRIZHEVSKY et al., 2012). Em relação a  $k = \{100, 150, 200\}$  podemos perceber que a diferença entre eles é relativamente pequena, porém quando se utiliza 100 blocos a abordagem CNN possui melhor desempenho, entretanto, na abordagem de Dissimilaridade a utilização de 150 blocos se destaca. Neste caso, decidimos que a melhor configuração é a divisão em 100 blocos pois possui melhor desempenho na abordagem da CNN, que é a parte mais custosa em relação ao desenvolvimento de todo método pelo fato da existência da limitação do *hardware*.

De acordo com os três experimentos realizados anteriormente, conseguimos chegar a seguinte configuração: a utilização de 2048 neurônios na penúltima camada, e a divisão de cada amostra em 100 blocos de tamanho  $128 \times 128$ . Assim, os próximos experimentos utilizarão esta configuração. A Tabela 5.2 apresenta detalhes da Figura 5.3 utilizando a base BFL com 115 escritores.

Os próximos experimentos serão realizados com as bases CVL e QUWI, respectivamente, que representam maior dificuldade na identificação de escritores, pois as duas



**Tabela 5.2.** Melhores Taxas de Identificação utilizando a Base BFL.

Quantidade de Blocos	Taxa de Identificação (%) (aceito para publicação)		
	CNN	Um-Contra-Todos	Dissimilaridade
100	95,65	90,43	96,52
150	93,91	97,39	98,26

apresentam amostras escritas em diferentes línguas pelo mesmo escritor. A CVL possui amostras em inglês e alemão, visto na Figura 5.4, porém as duas línguas utilizam alfabeto romano. Enquanto, a base QUWI possui amostras em inglês e árabe, que utilizam alfabetos diferentes aumentando a dificuldade de identificação, como podemos analisar na Figura 5.5.

Maitiffel is an Austrian nickname for the first computer working ~~to~~ solely on transistor on the European mainland. It was built in 1955 at the Vienna University of Technology by Heinz Zemanek. The builder plays on a quote on an operating computer: "If it is not the rapid calculation speed of American models that Whirlwind or Typhoon can achieve, it will be enough for a Wiener Maitiffel". The full name is Bitair dezimale Volltransistor-Rechenautomat.

(a)

Ward' ich zum Augenblicke sagen:  
Verweile doch! du bist so schön!  
Dann magst du mich in Fesek schlagen,  
Dann will ich gern zu Grunde gehn!  
Dann mag die Totkuglocke schallen,  
Dann bist du deines Diensts frey,  
Die Uhr mag stehn, der Zeiger fallen,  
Es sey die Zeit für mich vorbey!

(b)

**Figura 5.4.** Amostra (a) escrita em inglês e amostra (b) escrita em alemão do mesmo escritor da base CVL.

In a world where the beauty of Islam is shadowed by fear, the art speaks to the truth of this great way of life. Contemporary Muslim communities, regionally and around the globe, are facing a defining moment. The post 9/11 incident, as agreed upon by different researchers, established a new milestone.

(a)

وهل كان الهاجس المادي والتمويهات طاغية كما تقديركم للأعمال الإذاعية والتلفزيونية؟ جالفس كذا فقط نأخذ روايتنا، ومثل ذلك نذكر في هذا الامر لكوننا قنوعين جداً، وكنا نعتبر هذه الأشهر كأخوة لا نحن ها جسنا هو أرفيد المصنوع. وماذا نحن نواجه المذيع المحامي في التلفزيون؟ اليوم لو سألتك مثلاً عن عدد المذيعات القطريات في التلفزيون ماذا سيكون اجوابه؟ أكيد أن صناعة تراجم كبيراً للمذيع القطري.

(b)

**Figura 5.5.** Amostra (a) escrita em inglês e amostra (b) escrita em árabe do mesmo escritor da base QUWI.

Neste próximo experimento utilizamos a base CVL com 115 escritores, na qual cada amostra será dividida em 100 blocos de tamanho  $128 \times 128$  e 2048 neurônios na penúltima camada, ou seja, a quantidade de características. Diferente dos experimentos realizados anteriormente o conjunto de treinamento e o conjunto de teste será definido para cada execução, pois a base CVL possui amostras de diferentes línguas por escritor.

Desta forma, o intuito avaliar quando se utilizava todas as amostras, sendo uma para teste e as demais para treinamento, e quando utilizava apenas uma amostra no teste e no treinamento. As amostras escolhidas para serem utilizadas no conjunto de teste foram a terceira e a quinta amostra. A quinta amostra foi escrita em alemão e era diferente das demais, e a terceira, foi escrita em inglês e era a amostra que possuía uma quantidade média de conteúdo em relação as outras que também foram escritas em inglês, visto que a CVL possui apenas amostras com texto-dependente. A Tabela 5.3 apresenta as taxas de identificação obtidas utilizando a base CVL.

Os resultados apresentados através da Tabela 5.3, nos mostra que os dois primeiros experimentos empregam todas as amostras, sendo utilizado uma amostra para realizar o teste e o restante no conjunto de treinamento. A utilização das amostras em inglês (1,2,4) e a amostra em alemão (5) no conjunto de treinamento apresentou a melhor taxa de identificação, devido ao fato de possuir amostras da mesma língua no conjunto de treinamento e teste.

**Tabela 5.3.** Taxa de Identificação utilizando a base CVL.

Língua (Carta)		Taxa de Identificação (%)		
Treino	Teste	CNN	Um-Contra-Todos	Dissimilaridade
Inglês (1,2,4) e Alemão (5)	Inglês (3)	97,39	96,52	97,91
Inglês (1,2,3,4)	Alemão (5)	91,30	85,21	88,70
Alemão (5)	Inglês (3)	54,78	38,26	60,00
Inglês (3)	Alemão (5)	65,21	46,95	69,57

Enquanto, nos dois últimos experimentos o conjunto de treinamento possui um amostras em uma língua enquanto conjunto de teste possui amostras em outro idioma. Neste caso, ao usar amostras de manuscritos na língua inglesa para o treinamento e e testar em amostras em alemão, alcançamos melhores desempenhos. Isso pode acontecer pois o modelo gerado é melhor usando a língua inglesa, possivelmente por ser a língua nativa das escritores ou também pelo quantidade de texto cedido.

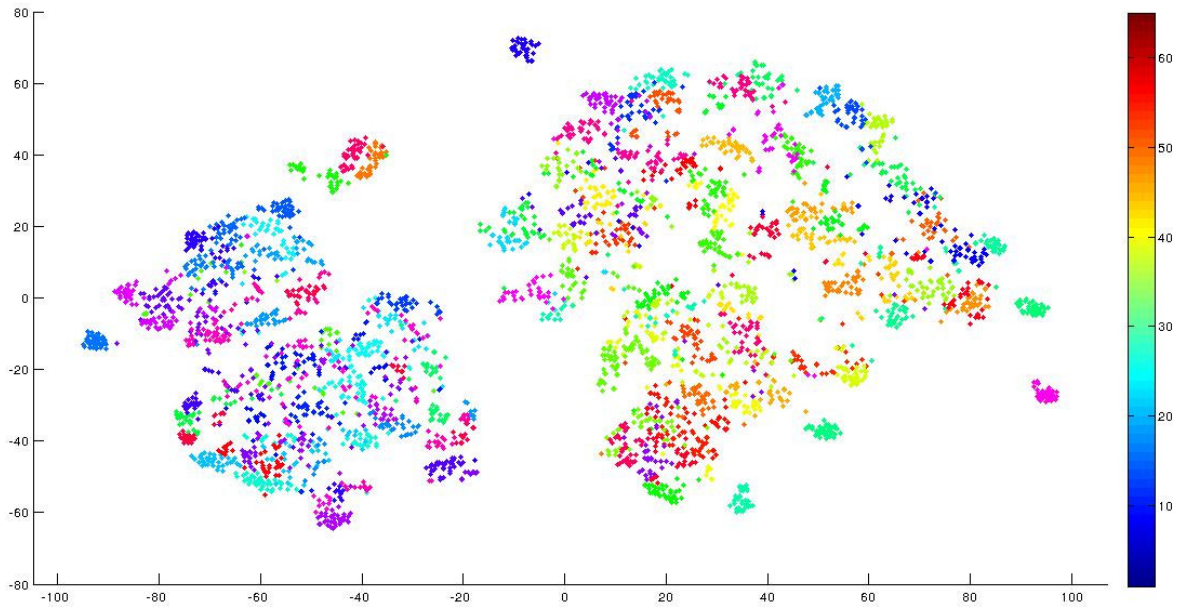
Um aspecto investigado neste trabalho e não levado em consideração em outros é o fato de avaliarmos as taxas utilizando uma língua diferente no conjunto de teste. O fato de empregar o mesmo alfabeto nos leva a crer que não haverá diferença nas taxas de identificação. Porém, avaliando as taxas, podemos concluir que ao empregar um estilo de escrita diferente, tivemos uma queda entre 6 e 11 pontos percentuais. Isso pode acontecer pois a as pessoas podem ter uma língua nativa e outra a qual não possui muita fluência, impactando diretamente no manuscrito.

A Figura 5.6 apresenta a distribuição do conjunto de treinamento do primeiro experimento, da abordagem Um-Contra-Todos, que utilizou a ferramenta *t-Distributed Stochastic Neighbor Embedding* (t-SNE) (MAATEN; HINTON, 2008). Podemos analisar que as cores representam cada escritor e em alguns casos as cores estão sobreposta o que significa maior dificuldade para o classificador distinguir cada classe.

No quinto experimento será utilizado a base QUWI com 115 escritores a qual cada amostra será dividida em 100 blocos de tamanho 128 e o conjunto de treinamento e teste será definido para cada configuração. A configuração utilizada entre conjunto de treinamento e teste pode ser dividida em duas partes, as quatro primeiras possuem amostras da mesma língua no treino e no teste, e as quatro últimas não utilizam amostras da mesma língua. Além disso, esta base possui uma diferença em relação as demais, pois ela possui duas amostras de texto-independente (Árabe (1) e Inglês (3)) e duas amostras de texto-dependente (Árabe (2) e Inglês (4)), portanto, foi utilizado todas as configurações possíveis. A Tabela 5.4 apresenta os resultados obtidos.

Analisando os resultados apresentados pela Tabela 5.4, podemos dividir entre os experimentos que utilizam amostras da mesma língua no conjunto de treinamento e teste e os experimentos que utilizam amostras de línguas distintas.

Em relação aos quatro primeiros experimentos, podemos analisar que utilizar amostras de texto-dependente no conjunto de treinamento geralmente apresentam melhor taxa de



**Figura 5.6.** Distribuição das amostras em inglês (1,2,4) e alemão (5) utilizando 115 escritores da base CVL.

**Tabela 5.4.** Taxa de Identificação utilizando a base QUWI.

Língua (Carta)		Taxa de Identificação (%)		
Treino	Teste	CNN	Um-Contra-Todos	Dissimilaridade
Inglês (3,4) e Árabe (2)	Árabe (1)	46,08	20,86	77,39
Inglês (3,4) e Árabe (1)	Árabe (2)	33,04	33,02	61,74
Árabe (1,2) e Inglês (4)	Inglês (3)	35,65	21,73	86,09
Árabe (1,2) e Inglês (3)	Inglês (4)	28,69	23,47	78,26
Inglês (3,4)	Árabe (1)	9,56	4,34	50,43
Inglês (3,4)	Árabe (2)	12,17	6,95	42,61
Árabe (1,2)	Inglês (3)	14,78	10,43	73,91
Árabe (1,2)	Inglês (4)	16,54	7,82	74,78

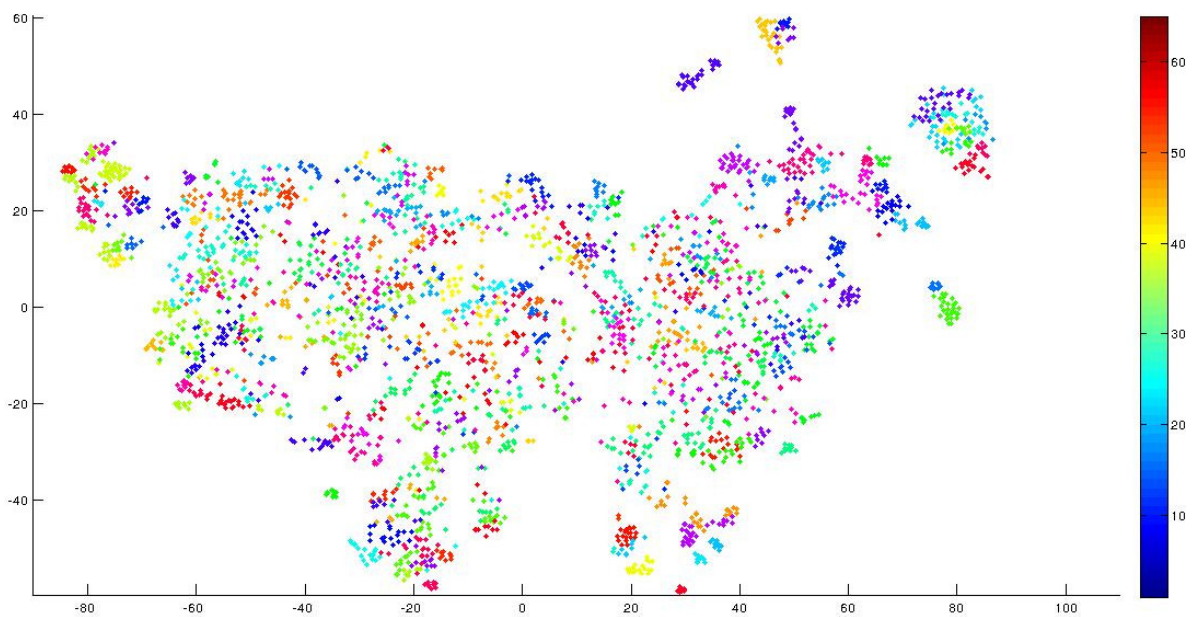
identificação. Provavelmente isto acontece pois a quantidade de informação presente em amostras de texto-dependente sempre é mesma para todos os escritores, enquanto para texto-independente a quantidade de informação pode variar, conseqüentemente, não sendo de grande utilidade para o conjunto de treinamento.

Nos quatro últimos experimentos podemos notar a taxa de identificação é maior quando se utiliza as amostras escritas em árabe (1,2) no conjunto de treinamento, enquanto, a taxa de identificação é bem menor quando se utiliza as amostras em inglês (3,4) no conjunto de treinamento. Portanto, podemos observar que a utilização de línguas diferentes no treino e no teste não sofre tanta influência em relação as amostras de texto-dependente e texto-independente como visto nos experimentos anteriores.

Para base QUWI, podemos perceber que as taxas são bem mais baixas se comparadas aos experimentos com a BFL e CVL. Isso nos mostra o quão desafiadora é esta base de dados. Ao possuir pelo menos uma amostra do mesmo estilo de escrita nos conjuntos de treinamento

e teste (primeira parte da tabela) é notável que as taxas são bem melhores. Entretanto, a dificuldade aumenta quando utilizamos um estilo para o treinamento e outro para o teste. Neste caso as taxas apresentadas nas abordagens CNN e Um-Contra-Todos são extremamente baixas. A abordagem de dissimilaridade apresenta uma boa vantagem sobre as demais, isso se deve ao método de escritor-independente, na qual as classes utilizadas no conjunto de teste não precisam estar no treinamento, basta existir um padrão significativo entre as instâncias das classes de teste.

A Figura 5.7 apresenta a distribuição do conjunto de treinamento do terceiro experimento, da abordagem Um-Contra-Todos, que também utilizou a t-SNE (MAATEN; HINTON, 2008), na qual, podemos analisar que as amostras estão bem mais distribuídas em relação a Figura 5.6, ou seja, existe maior dificuldade em realizar a classificação da base QUWI.



**Figura 5.7.** Distribuição das amostras em árabe (1,2) e inglês (4) utilizando 115 escritores da base QUWI.

Nos próximos experimentos utilizamos os melhores resultados obtidos de cada base, e suas respectivas configurações, porém as características foram extraídas de outros 115 escritores, visto que todas as bases possuem mais que 230 escritores. Além disso, as características foram extraídas utilizando os pesos dos respectivos experimentos anteriores de cada base, ou seja, criamos o modelo com os pesos já extraídos e através do modelo, extraímos as características dos escritores que não foram utilizados pela CNN, desta forma, os resultados obtidos abrangem somente as abordagens Um-Contra-Todos e Dissimilaridade. A Tabela 5.5 apresenta os resultados obtidos de cada base.

A partir dos resultados apresentados pela Tabela 5.5, podemos notar que na maioria dos casos a taxa de identificação se mantém em comparação com os experimentos utilizados

**Tabela 5.5.** Taxa de identificação utilizando outros 115 escritores de cada base.

Base	Língua (Carta)		Taxa de Identificação (%)	
	Treino	Teste	Um-Contra-Todos	Dissimilaridade
BFL	Português (1,2)	Português (3)	93,04	98,26
CVL	Inglês (1,2,4) e Alemão (5)	Inglês (3)	86,08	93,74
CVL	Inglês (1,2,3,4)	Alemão (5)	72,06	84,13
QUWI	Inglês (3,4) e Árabe (2)	Árabe (1)	45,21	71,69
QUWI	Árabe (1,2) e Inglês (4)	Inglês (3)	37,39	86,96

anteriormente. Portanto, podemos analisar que as características extraídas pelo modelo da CNN são robustas mesmo que os pesos utilizados sejam de outros escritores de cada base.

Por fim, foi realizado um último experimento buscando avaliar a base toda. Neste caso utilizamos somente duas abordagens, Um-Contra-Todos e Dissimilaridade. Na abordagem Um-Contra-Todos, extraímos as características de todas as amostras da mesma forma que foi realizado no experimento anterior. Entretanto, para a Dissimilaridade, utilizamos os modelos gerados por 115 escritores para classificar a base toda. Isto é possível pois a dissimilaridade realiza uma transformação para o espaço de dissimilaridade, passando a trabalhar com modelos binários, desta forma não é possível inserir novos escritores sem a necessidade de retrainar o modelo. A Tabela 5.6 apresenta os resultados obtidos de cada base utilizando todos os seus respectivos escritores, sendo BFL com 315, CVL com 309 e QUWI com 475.

**Tabela 5.6.** Taxa de identificação utilizando todos escritores de cada base.

Base	Língua (Carta)		Taxa de Identificação (%)	
	Treino	Teste	Um-Contra-Todos	Dissimilaridade
BFL	Português (1,2)	Português (3)	86,34	91,56
CVL	Inglês (1,2,4) e Alemão (5)	Inglês (3)	79,57	87,82
CVL	Inglês (1,2,3,4)	Alemão (5)	68,22	79,61
QUWI	Inglês (3,4) e Árabe (2)	Árabe (1)	25,26	62,79
QUWI	Árabe (1,2) e Inglês (4)	Inglês (3)	23,32	77,60

Analisando os resultados apresentados pela Tabela 5.6, podemos observar que em geral a taxa de identificação é menor em comparação com a Tabela 5.5. Porém em relação a abordagem Dissimilaridade conseguimos afirmar que o modelo gerado por 115 escritores se mantém robusto mesmo com o aumento do número de escritores no conjunto de teste. Entretanto, não conseguimos afirmar se o modelo se mantém robusto, ao aumentarmos o número de escritores no conjunto de teste mais do que utilizado neste experimento, ou se utilizarmos menos escritores no conjunto de treinamento para gerar o modelo.

---

## Conclusão

---

Neste capítulo apresentaremos a conclusão a partir dos resultados obtidos nos experimentos, considerando as abordagens CNN, Um-Contra-Todos e Dissimilaridade.

O principal objetivo deste trabalho foi avaliar as taxas de desempenho na identificação de escritores utilizando a técnica de geração de textura considerando três bases de dados. O método proposto consistiu na aplicação da CNN, com a função de realizar a classificação e a extração de características, que posteriormente foram utilizadas para a classificação no SVM, e na utilização da abordagem de dissimilaridade que também utilizou o classificador SVM. Além disso, algumas avaliações específicas foram realizadas como a análise do número de características extraídas pela CNN e a quantidade e tamanho dos blocos utilizado no conjunto de treinamento e teste. Este trabalho se distingue pelo uso da abordagem de dissimilaridade nas características extraídas pela CNN aplicado ao problema de identificação de escritores.

A partir dos resultados obtidos, é possível concluir que a utilização da abordagem de Dissimilaridade apresenta melhores resultados se comparado as abordagens Um-Contra-Todos e CNN. Comparando com os trabalhos relacionados visto na Tabela 2.1, as taxas alcançadas são melhores que neste trabalho, porém em alguns casos os trabalhos relacionados não realizam a identificação separando o conjunto de treinamento e teste pela língua que cada base de dados possui. Assim, podemos concluir que apesar das taxas em alguns casos não serem maiores que as apresentadas em literatura, o modelo gerado pelo método utilizado no trabalho se mantém robusto para diferentes línguas presentes na base de dados.

### 6.1. Trabalhos Futuros

Durante este trabalho, observamos algumas questões que podem ser relevantes e podem ser investigadas futuramente.

- **Outras bases de dados:** Neste trabalho foram utilizados três bases de dados, sendo duas multi-script. Entretanto a utilização de outras bases de dados com outras línguas pode contribuir para avaliar o comportamento do método;
- **Outra arquitetura para CNN:** A arquitetura utilizada foi definida por experimentação, porém existem algumas arquiteturas definidas e bastante utilizadas por alguns trabalhos que podem ser utilizadas neste trabalho, desconsiderando a limitação de *hardware*;
- **Modelo gerado pela Dissimilariade:** Utilizamos o modelo gerado pelo SVM através de 115 escritores, e classificamos o resto dos escritores da base de dados. Entretanto, seria interessante avaliar se o modelo gerado obtém taxas boas se aumentarmos o número de escritores no conjunto de teste ou gerarmos o modelo do SVM utilizando menos escritores.



# Referências

---

- ABDEL-HAMID, Ossama et al. Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, Canada: IEEE, 2012. p. 4277–4280.
- AHONEN, Timo; HADID, Abdenour; PIETIKAINEN, Matti. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 28, n. 12, p. 2037–2041, 2006. ISSN 0162-8828.
- AL-MAADEED, Somaya et al. Novel geometric features for off-line writer identification. *Pattern Analysis and Applications*, Springer, p. 1–10, 2014.
- AMARAL, Aline Maria Malachini Miotto. *Identificação de autoria de documentos manuscritos utilizando características grafométricas*. Tese (Doutorado) — Pontifícia Universidade Católica do Paraná, 2014.
- AREL, Itamar; ROSE, Derek C; KARNOWSKI, Thomas P. Deep machine learning—a new frontier in artificial intelligence research. *IEEE Computational Intelligence Magazine*, IEEE, v. 5, n. 4, p. 13–18, 2010. ISSN 1556-603X.
- BERTOLINI, Diego et al. Texture-based descriptors for writer identification and verification. *Expert Systems with Applications*, Elsevier, v. 40, n. 6, p. 2069–2080, 2013.
- CHANG, Chih-Chung; LIN, Chih-Jen. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, v. 2, p. 27:1–27:27, 2011. Software available at <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>.
- CHOLLET, François. *Keras*. 2015. Programa de computador. Disponível em: <<https://github.com/fchollet/keras>>.
- CHRISTLEIN, Vincent et al. Offline writer identification using convolutional neural network activation features. In: *Pattern Recognition: 37th German Conference, GCPR 2015, Aachen, Germany, October 7-10, 2015, Proceedings*. Cham: Springer International Publishing, 2015. p. 540–552.
- DJEDDI, Chawki et al. Text-independent writer recognition using multi-script handwritten texts. *Pattern Recognition Letters*, Elsevier, v. 34, n. 10, p. 1196–1202, 2013.
- FIEL, Stefan; SABLATNIG, Robert. Writer identification and retrieval using a convolutional neural network. In: *Computer Analysis of Images and Patterns*. Cham: Springer, 2015. p. 26–37.
- FREITAS, C et al. Brazilian forensic letter database. In: *11th International workshop on frontiers on handwriting recognition*. Montreal, Canada: ICFHR, 2008.

- HAFEMANN, Luiz Gustavo. *An Analysis of Deep Neural Networks for Texture Classification*. Dissertação (Mestrado) — Universidade Federal do Paraná, 2014.
- HAFEMANN, Luiz G; SABOURIN, Robert; OLIVEIRA, Luiz S. Analyzing features learned for offline signature verification using deep cnns. *arXiv preprint arXiv:1607.04573*, 2016.
- HAFEMANN, Luiz G.; SABOURIN, Robert; OLIVEIRA, Luiz S. Writer-independent feature learning for offline signature verification using deep convolutional neural networks. *CoRR*, abs/1604.00974, 2016.
- HANNAD, Yaacoub; SIDDIQI, Imran; KETTANI, Mohamed El Youssfi El. Writer identification using texture descriptors of handwritten fragments. *Expert Systems with Applications*, Elsevier, v. 47, p. 14–22, 2016.
- HANUSIAK, R. K. et al. Writer verification using texture-based features. *International Journal on Document Analysis and Recognition (IJDAR)*, Springer, v. 15, n. 3, p. 213–226, 2011. ISSN 1433-2825.
- HASSAINE, Abdulaali et al. Icdar 2013 competition on gender prediction from handwriting. In: *12th International Conference on Document Analysis and Recognition*. Washington, DC: IEEE, 2013. p. 1417–1421. ISSN 1520-5363.
- HE, Sheng; SCHOMAKER, Lambert. Writer identification using curvature-free features. *Pattern Recognition*, Elsevier, v. 63, p. 451–464, 2016.
- HINTON, Geoffrey E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.
- JAIN, Anil K; DUIN, Robert PW; MAO, Jianchang. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 22, n. 1, p. 4–37, 2000. ISSN 0162-8828.
- JARRETT, Kevin et al. What is the best multi-stage architecture for object recognition? In: *IEEE 12th International Conference on Computer Vision*. New York, USA: IEEE, 2009. p. 2146–2153.
- KARN, Ujjwal. *An Intuitive Explanation of Convolutional Neural Networks*. 2016. Disponível em: <<https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>>.
- KITTLER, Josef et al. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 20, n. 3, p. 226–239, 1998.
- KLEBER, Florian et al. Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In: *12th International Conference on Document Analysis and Recognition*. Washington, DC: IEEE, 2013. p. 560–564. ISSN 1520-5363.
- KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. NIPS: Curran Associates, Inc., 2012. v. 25, p. 1097–1105.
- LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

- LORENA, Ana Carolina; CARVALHO, André CPLF de. Introduçãoas máquinas de vetores suporte. *Revista de Informática Teórica e Aplicada*, v. 14, p. 43–67, 2007. ISSN 0103-2569.
- MÁADEED, Somaya Al et al. Quwi: an arabic and english handwriting dataset for offline writer identification. In: *International Conference on Frontiers in Handwriting Recognition (ICFHR)*. Doha, Qatar: IEEE, 2012. p. 746–751.
- MAATEN, Laurens van der; HINTON, Geoffrey. Visualizing data using t-sne. *Journal of Machine Learning Research*, v. 9, n. Nov, p. 2579–2605, 2008.
- MARTINS, J. G. et al. Forest species recognition based on dynamic classifier selection and dissimilarity feature vector representation. *Machine Vision and Applications*, Springer, v. 26, n. 2, p. 279–293, 2015. ISSN 1432-1769.
- NIELSEN, Michael. Deep learning. In: *Neural Networks and Deep Learning*. Michael Nielsen, 2016. cap. 6. Disponível em: <<http://neuralnetworksanddeeplearning.com/chap6.html>>.
- OTSU, Nobuyuki. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, IEEE, v. 9, p. 62–66, 1979. ISSN 0018-9472.
- P’KALSKA, Elżbieta; DUIN, Robert PW. Classifiers for dissimilarity-based pattern recognition. In: *15th International Conference on Pattern Recognition, 2000. Proceedings*. The Netherlands: IEEE, 2000. v. 2, p. 12–16.
- SIDDIQI, Imran; VINCENT, Nicole. A set of chain code based features for writer recognition. In: *10th International Conference on Document Analysis and Recognition*. Paris, França: IEEE, 2009. p. 981–985.
- TRIER, Øivind Due; JAIN, Anil K; TAXT, Torfinn. Feature extraction methods for character recognition-a survey. *Pattern recognition*, Elsevier, v. 29, n. 4, p. 641–662, 1996.
- WARD, Jonatan et al. Efficient mapping of the training of convolutional neural networks to a cuda-based cluster. Eindhoven University of Technology, 2011.