

***VITOR ARINS PINTO***

***REDES NEURAIAS CONVOLUCIONAIS DE PROFUNDIDADE PARA  
RECONHECIMENTO DE TEXTOS EM IMAGENS DE CAPTCHA.***

Florianópolis

21 de julho de 2016

**VITOR ARINS PINTO**

***REDES NEURAIS CONVOLUCIONAIS DE PROFUNDIDADE PARA  
RECONHECIMENTO DE TEXTOS EM IMAGENS DE CAPTCHA.***

Trabalho de Conclusão de Curso submetido ao  
Programa de graduação da Universidade Federal de Santa Catarina para a obtenção do Grau de Bacharel em Sistemas de Informação.

Orientadora: Luciana de Oliveira Rech

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA

Florianópolis

21 de julho de 2016

## ***AGRADECIMENTOS***

## ***RESUMO***

Atualmente muitas aplicações na Internet seguem a política de manter alguns dados acessíveis ao público. Para isso é necessário desenvolver um portal que seja robusto o suficiente para garantir que todas as pessoas possam acessá-lo. Porém as requisições feitas para recuperar dados públicos nem sempre vêm de um ser humano. Empresas especializadas em Big data possuem um grande interesse em fontes de dados públicos para poder fazer análises e previsões a partir de dados atuais. Com esse interesse, Web Crawlers são implementados. Eles são responsáveis por consultar fontes de dados milhares de vezes ao dia, fazendo diversas requisições a um site. Tal site pode não estar preparado para um volume de consultas em um período tão curto de tempo. Com o intuito de impedir que sejam feitas consultas por programas de computador, as instituições que mantêm dados públicos investem em ferramentas chamadas CAPTCHA (teste de Turing público completamente automatizado, para diferenciação entre computadores e humanos). Essas ferramentas geralmente se tratam de imagens contendo um texto qualquer e o usuário deve digitar o que vê na imagem. O objetivo do trabalho proposto é realizar o reconhecimento de texto em imagens de CAPTCHA através da aplicação de redes neurais convolucionais.

## ***ABSTRACT***

Currently many applications on the Internet follow the policy of keeping some data accessible to the public. For this it is necessary to develop a portal that is robust enough to ensure that all people can access this data. But the requests made to recover Public Data not always come from a human. companies specializing in Big data have a great interest in data from public sources in order to make analyzes and forecasts from current data. With this interest, Web Crawlers are implemented. They are responsible for querying data sources thousands of times a day, making several requests to a site. This site may not be prepared for a volume of inquiries in a short period of time. In order to prevent queries to be made by computer programs, institutions that keep public data invest in tools called CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*). These tools usually deal with images containing text and the user must enter what he or she sees in the image. The objective of the proposed work is to perform the text recognition in CAPTCHA images through the application of convolutional neural networks.

## ***LISTA DE FIGURAS***

Figura 1	Aplicação de uma função “kernel” sobre a função de uma imagem que é o “input”. .....	5
----------	--	---

## ***LISTA DE TABELAS***

## ***LISTA DE ABREVIATURAS E SIGLAS***

CAPTCHA . . . . . *Completely Automated Public Turing test to tell Computers and Humans Apart*



## ***SUMÁRIO***

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>1</b>
1.1	PROBLEMA .....	1
1.2	OBJETIVOS .....	2
1.2.1	<i>OBJETIVO GERAL</i> .....	2
1.2.2	<i>OBJETIVOS ESPECÍFICOS</i> .....	2
1.3	ESCOPO DO TRABALHO .....	2
1.4	METODOLOGIA .....	2
1.5	ESTRUTURA DO TRABALHO .....	2
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>3</b>
2.1	APRENDIZADO DE MÁQUINA .....	3
2.2	REDES NEURAIS .....	3
2.3	CONVOLUÇÕES .....	4
2.4	APRENDIZADO EM PROFUNDIDADE .....	5
2.5	REDES NEURAIS CONVOLUCIONAIS DE PROFUNDIDADE .....	5
	<b>REFERÊNCIAS .....</b>	<b>6</b>

# 1 INTRODUÇÃO

Redes neurais artificiais clássicas existem desde os anos 60, como fórmulas matemáticas e algoritmos. Atualmente os programas de aprendizado de máquina contam com diferentes tipos de redes neurais. Um tipo de rede neural muito utilizado para processamento de imagens é a rede neural convolucional de profundidade. O trabalho em questão tratará da utilização e configuração de uma rede neural convolucional de profundidade para reconhecimento de textos em imagens.

## 1.1 PROBLEMA

Com o aumento constante na quantidade de informações geradas e computadas atualmente, percebe-se o surgimento de uma necessidade de tornar alguns tipos de dados acessíveis a um público maior. A fim de gerar conhecimento, muitas instituições desenvolvem portais de acesso para consulta de dados relevantes a cada pessoa. Esses portais, em forma de aplicações na Internet, precisam estar preparados para receber diversas requisições e em diferentes volumes ao longo do tempo.

Devido a popularização de ferramentas e aplicações especializadas em Big data, empresas de tecnologia demonstram interesse em recuperar grandes volumes de dados de diferentes fontes públicas. Para a captura de tais dados, Web crawlers são geralmente implementados para a realização de várias consultas em aplicações que disponibilizam dados públicos.

Para tentar manter a integridade da aplicação, as organizações que possuem estas informações requisitadas investem em ferramentas chamadas CAPTCHA (teste de Turing público completamente automatizado para diferenciação entre computadores e humanos). Essas ferramentas frequentemente se tratam de imagens contendo um texto qualquer e o usuário precisa digitar o que vê na imagem.

O trabalho de conclusão de curso proposto tem a intenção de retratar a ineficiência de algumas ferramentas de CAPTCHA, mostrando como redes neurais convolucionais podem ser

aplicadas em imagens a fim de reconhecer o texto contido nestas imagens.

## 1.2 OBJETIVOS

### 1.2.1 *OBJETIVO GERAL*

Analisar o treinamento e aplicação de redes neurais convolucionais de profundidade para o reconhecimento de texto em imagens de CAPTCHA.

### 1.2.2 *OBJETIVOS ESPECÍFICOS*

- Estudar trabalhos correlatos e analisar o estado da arte;
- Entender como funciona cada aspecto na configuração de uma rede neural convolucional;
- Realizar o treinamento e aplicação de uma rede neural artificial para reconhecimento de CAPTCHAs.

## 1.3 ESCOPO DO TRABALHO

O escopo deste trabalho inclui o estudo e análise de redes neurais convolucionais de profundidade para reconhecimento de texto em imagens de CAPTCHA. Não está no escopo do trabalho analisar outras formas de inteligência no reconhecimento de texto. Também não está no escopo do trabalho o estudo, análise ou implementação da aplicação de redes neurais convolucionais para outros tipos de problemas.

## 1.4 METODOLOGIA

Para realizar o proposto, serão feitas pesquisas em base de dados tais como IEE Xplorer , ACM Portal, adquirindo assim maior conhecimento sobre o tema, estudando trabalhos relacionados sobre o tema.

Com o estado da arte já estudado e conhecido, serão feitas pesquisas e estudos para indicar caminhos possíveis para desenvolvimento da proposta de TCC.

## 1.5 ESTRUTURA DO TRABALHO

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 APRENDIZADO DE MÁQUINA

Aprendizado de máquina, ou *Machine Learning*, é uma área da computação que emergiu de estudos relacionados ao reconhecimento de padrões e inteligência artificial. Nesta área é contemplado o estudo e implementação de algoritmos que conseguem aprender e fazer previsões baseadas em dados. Esses algoritmos funcionam através da construção de um modelo preditivo que tem como entrada um conjunto de treinamento com dados de observações quaisquer. Desse modo as previsões são feitas orientadas aos dados e não a partir de instruções estáticas de um programa.

### 2.2 REDES NEURAIS

Diante das ferramentas disponíveis que tratam de aprendizado de máquina, uma delas é a rede neural artificial.

Redes neurais artificiais são conjuntos de modelos inspirados por redes neurais biológicas, usados para aproximar funções que dependem de um número muito grande de entradas. De acordo com Mackay[1], Redes neurais geralmente são especificadas utilizando 3 coisas:

- **Arquitetura:** Especifica quais variáveis estão envolvidas na rede e quais as relações topológicas. Por exemplo, as variáveis envolvidas em uma rede neural podem ser os pesos das conexões entre os neurônios.
- **Regra de atividade:** A maioria dos modelos de rede neural tem uma dinâmica de atividade com escala de tempo curta. São regras locais que definem como as *atividades* de neurônios mudam em resposta aos outros. Geralmente a regra de atividade depende dos parâmetros da rede.
- **Regra de aprendizado:** Especifica o modo com que os pesos da rede neural muda conforme o tempo. O aprendizado normalmente toma uma escala de tempo maior do que

a escala referente a dinâmica de atividade. Normalmente a regra de aprendizado dependerá das *atividades* dos neurônios. Também pode depender dos valores que são objetivos definidos pelo usuário e valores iniciais dos pesos.

Tomando imagens como exemplo, uma rede neural para reconhecimento de texto pode ter como entrada o conjunto de pixels da imagem. Depois de serem atribuídos os pesos para cada item da entrada, os próximos neurônios serão ativados mediante a função de atividade pré-definida. Os pesos são recalculados através da regra de aprendizado e todo processo é repetido até uma condição determinada pelo usuário.

## 2.3 CONVOLUÇÕES

Para entender redes neurais convolucionais, é necessário primeiro entender o que são convoluções. Segundo Olah[2], uma convolução pode ser vista como um somatório das probabilidades de resposta de duas funções algébricas. Tendo como definição padrão de convolução a seguinte expressão:

$$(f * g)(c) = \sum_a f(a) \cdot g(c - a) \quad (2.1)$$

Onde  $f$  e  $g$  são duas funções,  $c$  é o parâmetro de entrada para a função final e  $a$  é um parâmetro de entrada escolhido para uma das funções, geralmente uma diferença temporal.

Como podemos considerar que imagens são funções bidimensionais, é comum realizar transformações por meio de convoluções. Estas convoluções são executadas com uma função local pequena chamada de “kernel”.

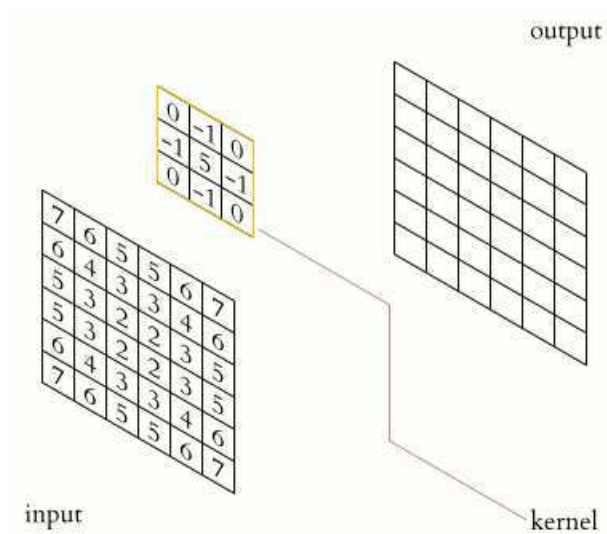


Figura 1: Aplicação de uma função “kernel” sobre a função de uma imagem que é o “input”.

## 2.4 APRENDIZADO EM PROFUNDIDADE

Citar Hinton[3] e Jaderberg[4]

## 2.5 REDES NEURASIS CONVOLUCIONAIS DE PROFUNDIDADE

Citar Krizhevsky[5] e Goodfellow[6]

## **REFERÊNCIAS**

- [1] MACKAY, D. J. C. *Information Theory , Inference And Learning Algorithms*. Cambridge University Press, 2005. ISBN 9780521670517. Disponível em: <<http://www.inference.phy.cam.ac.uk/mackay/itila/>>.
- [2] OLAH, C. *Understanding Convolutions*. 2014. Acessado: 18/07/2016. Disponível em: <<http://colah.github.io/posts/2014-07-Understanding-Convolutions/>>.
- [3] HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A fast learning algorithm for deep belief nets. Disponível em: <<https://www.cs.toronto.edu/hinton/absps/fastnc.pdf>>.
- [4] JADERBERG, M. et al. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition. Disponível em: <<http://arxiv.org/pdf/1406.2227v4.pdf>>.
- [5] KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks.
- [6] GOODFELLOW, I. J. et al. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. Disponível em: <<http://arxiv.org/pdf/1312.6082v4.pdf>>.