

SEMINÁRIO DOS ARTIGOS:

**Text Detection and Character Recognition in
Scene Images with Unsupervised Feature
Learning**

**End-to-End Text Recognition with Convolutional
Neural Networks**



Fernanda Maria

Sirlene

SUMÁRIO

- Introdução
- Learning Architecture
- 1º artigo: Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning
- 2º artigo: End-to-End Text Recognition with Convolutional Neural Networks
- Resultados e conclusões
- Referências

INTRODUÇÃO

- Detecção de texto e identificação de caracteres em imagens naturais é um problema de reconhecimento visual desafiador.



- Ampla gama de variações nas condições de fundos, texturas, fontes e iluminação.



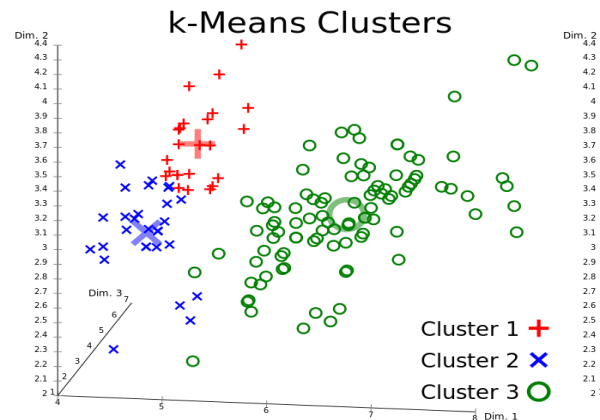
INTRODUÇÃO

- Diversas propostas e algoritmos na área geralmente com foco específico
- Modelos probabilístico:
 - Sistema de detecção
 - Sistemas de reconhecimento.
- Sistemas de aprendizagem altamente flexíveis:
 - Arquiteturas de rede neural multicamadas

LEARNING ARCHITECTURE

- K-means clustering

- Técnica que usa o algoritmo de agrupamento de dados por K-médias.



- Objetivo: encontrar a melhor divisão de P dados em K grupos C_i , $i = 1, \dots, K$, de maneira que a distância total entre os dados de um grupo e o seu respectivo centro, somada por todos os grupos, seja minimizada.

LEARNING ARCHITECTURE

- Rede Neural Covolucional(CNN)
 - Múltiplas camadas constituídas por diferentes fins.
 - Primeira:
 - Fase de extração de características, que consiste em neurônios convolucionais e redução da resolução.
 - Final:
 - Neurônios da rede perceptron realizam a classificação das características extraídas.

TEXT DETECTION AND CHARACTER RECOGNITION IN SCENE IMAGES WITH UNSUPERVISED FEATURE LEARNING

- Algoritmos de grande escala para aprender os recursos automaticamente a partir de dados sem definições.
- Mostrar classificadores altamente eficazes tanto para a detecção e reconhecimento para ser usado em uma alta precisão sistema end-to-end.

TEXT DETECTION AND CHARACTER RECOGNITION IN SCENE IMAGES WITH UNSUPERVISED FEATURE LEARNING

1. Coletar um conjunto de patches pequena imagem, $x^{(i)}$ de formação dados. Escala de cinza 8x8 patches.
2. Aplicar simples pré-tratamento estatístico (clareamento) para as porções de entrada para produzir um novo conjunto de dados $x^{(i)}$.
3. Executar um algoritmo de aprendizado não supervisionado nas $x^{(i)}$ a construir um mapeamento de fragmentos de entrada para um vetor de características, $z^{(i)} = f(x^{(i)})$

TEXT DETECTION AND CHARACTER RECOGNITION IN SCENE IMAGES WITH UNSUPERVISED FEATURE LEARNING

- Aprendizagem de Características
 - Dado um conjunto de imagens de treinamento:
 - Extraíram um conjunto de 8 por 8 pixels para produzir vetores.
 - Branqueamento.
 - Estágio de aprendizado não supervisionado:
 - Variante do K-means clustering.
 - Aprende-se um conjunto de vetores normalizados $D^{(i)}$, $j \in \{1, \dots, d\}$ para formar as colunas de D , utilizando produto interno como semelhança métrica.
 - A otimização é feita por minimização sobre D e os $s^{(i)}$ alternadamente.

TEXT DETECTION AND CHARACTER RECOGNITION IN SCENE IMAGES WITH UNSUPERVISED FEATURE LEARNING

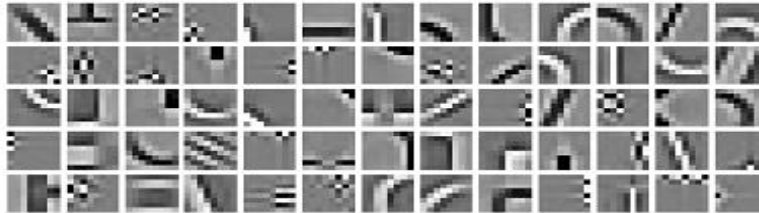


Figure 1. A small subset of the dictionary elements learned from grayscale, 8-by-8 pixel image patches extracted from the ICDAR 2003 dataset.

- Dicionário treinado:
 - Definir as características para um único novo patch de 8 por 8.
- Dado um novo patch de entrada x :
 - Aplicar a normalização e branqueamento já utilizado.
 - Mapeá-lo para uma nova representação $z \in \mathbb{R}^d$ tomando o produto interno com cada elemento do dicionário e aplicar uma função escalar não-linear.

$$z = \max \{0, |D \cdot x| - \alpha\}$$

TEXT DETECTION AND CHARACTER RECOGNITION IN SCENE IMAGES WITH UNSUPERVISED FEATURE LEARNING

○ Extração de características

- Reduz a dimensionalidade da representação antes da classificação.
 - Pooling média: soma os vetores $z^{(i,j)}$ de mais de 9 blocos em uma grade 3-por-3 sobre a imagem, gerando um vetor de características final com características 9d para esta imagem

○ Treinamento do detector de texto

- Treinou-se um classificador binário que visa distinguir as que possuem texto ou não.
- Conjunto de treinamento:
 - Windows de 32-by-32 a partir do conjunto de dados Treinamento ICDAR 2003.

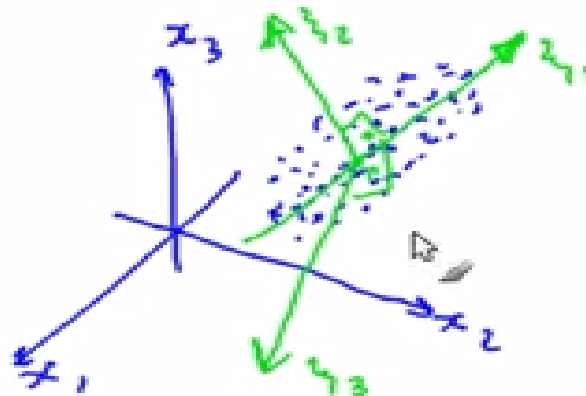
ETAPAS DO SISTEMA DESCRITO EM: END-TO-END TEXT RECOGNITION WITH CONVOLUTIONAL NEURAL NETWORKS

- 1) Construção do banco de características (filtros);
- 2) Avaliação das características convolucionalmente e redução do número de características;
- 3) Treinamento de classificadores lineares:
 - Detecção de texto;
 - Reconhecimento de caracteres.
- 4) Integração com dicionário.

BANCO DE CARACTERÍSTICAS

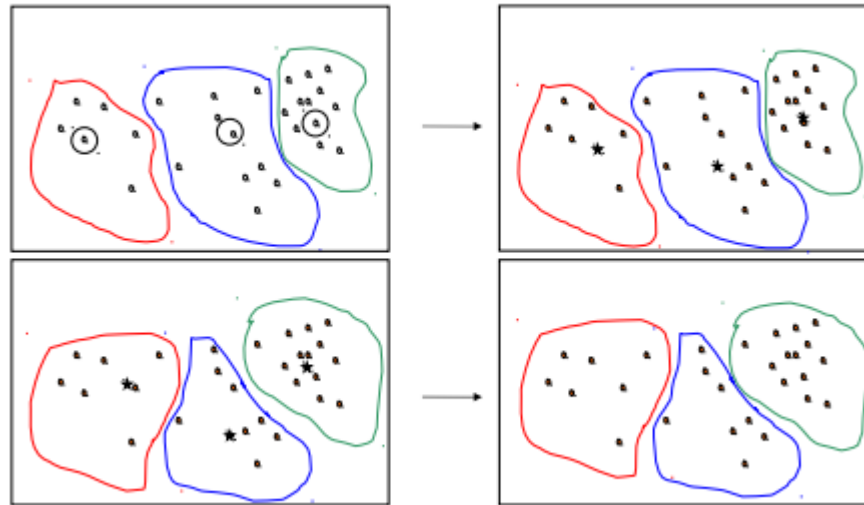
- Selecciona-se m blocos 8x8;
- Normalização:
 - Subtrai-se a média e divide-se pelo desvio padrão;
- Branqueamento:
 - ZCA: é como Análise de Componentes Principais (PCA), porém rotaciona os dados de volta para os mesmos eixos da entrada original.

Fig.: Exemplo PCA



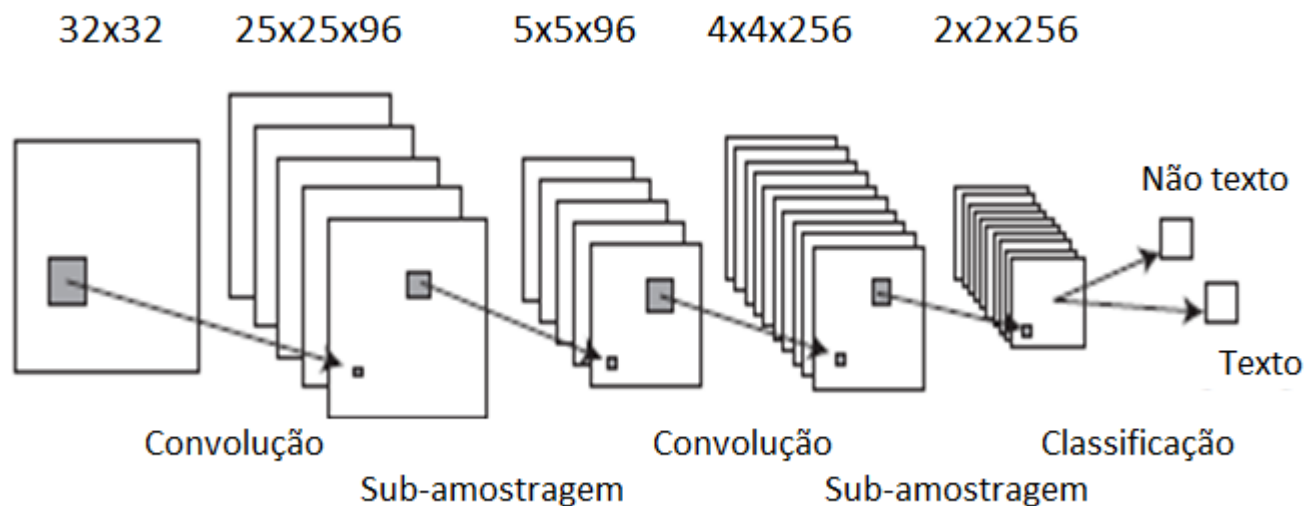
BANCO DE CARACTERÍSTICAS

- Variante K-means: banco de $n1$ filtros D ($64 \times n1$);



REDE NEURAL CONVOLUCIONAL

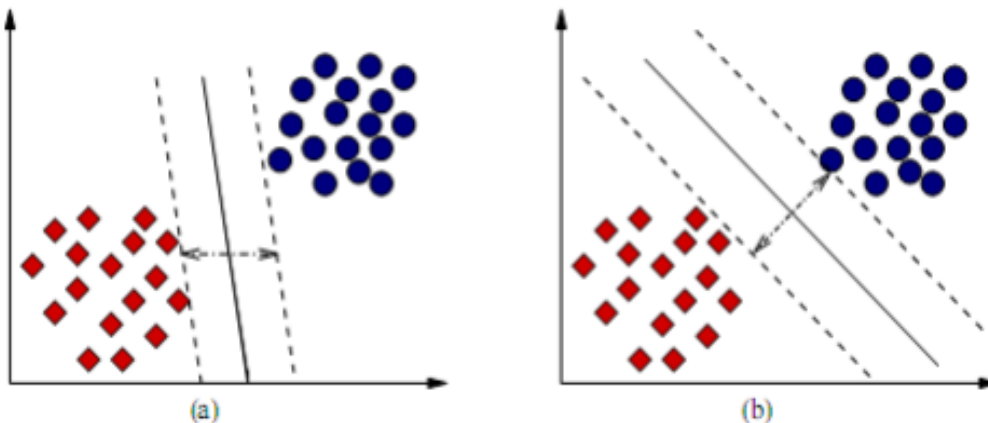
- Diante da imagem de entrada de 32x32, calculamos z para cada sub-janela 8x8 para obter o mapa de características 25x25x1 resultante da primeira camada.
- $z = \max \{0, |D \cdot x| - \alpha\}$, onde $\alpha = 0,5$



REDE NEURAL CONVOLUCIONAL

- Sub-amostragem: agrupamento médio;
- Classificação: SVM (Máquina de Vetores de Suporte)

Fig.: (a) Hiperplano com margem pequena. (b) Hiperplano com margem máxima.



INTEGRAÇÃO: MÓDULOS DE DETECÇÃO E RECONHECIMENTO COM UM DICIONÁRIO

- Aplica-se NMS (Non-Maximum Suppression) sobre as resposta do detector de texto para:
 - Identificar as linhas horizontais,
 - Remover caixas delimitadoras sobrepostas;
 - Detectar lacunas.
- Segmentação de caracteres: Beam Search;
- Utilizando o módulo de reconhecimento, cria uma matriz M de pontuação, $62 \times N$;
- Encontra-se a palavra do dicionário w^* que mais combina com uma matriz de pontuação M .

RESULTADOS E CONCLUSÕES

- Maiores bancos características conseguem aumentar a precisão [1];
- O sistema alcança o desempenho dos métodos estado da arte em conjuntos de teste padrão [2].

REFERÊNCIAS

- [1] Coates, A., Carpenter, B., Case, C., Satheesh, S., Suresh, B., Wang, T., ... & Ng, A. Y. (2011, September). Text detection and character recognition in scene images with unsupervised feature learning. In Document Analysis and Recognition (ICDAR), 2011 International Conference on (pp. 440-445). IEEE.
- [2] Wang, T., Wu, D. J., Coates, A., & Ng, A. Y. (2012, November). End-to-end text recognition with convolutional neural networks. In Pattern Recognition (ICPR), 2012 21st International Conference on (pp. 3304-3308). IEEE.