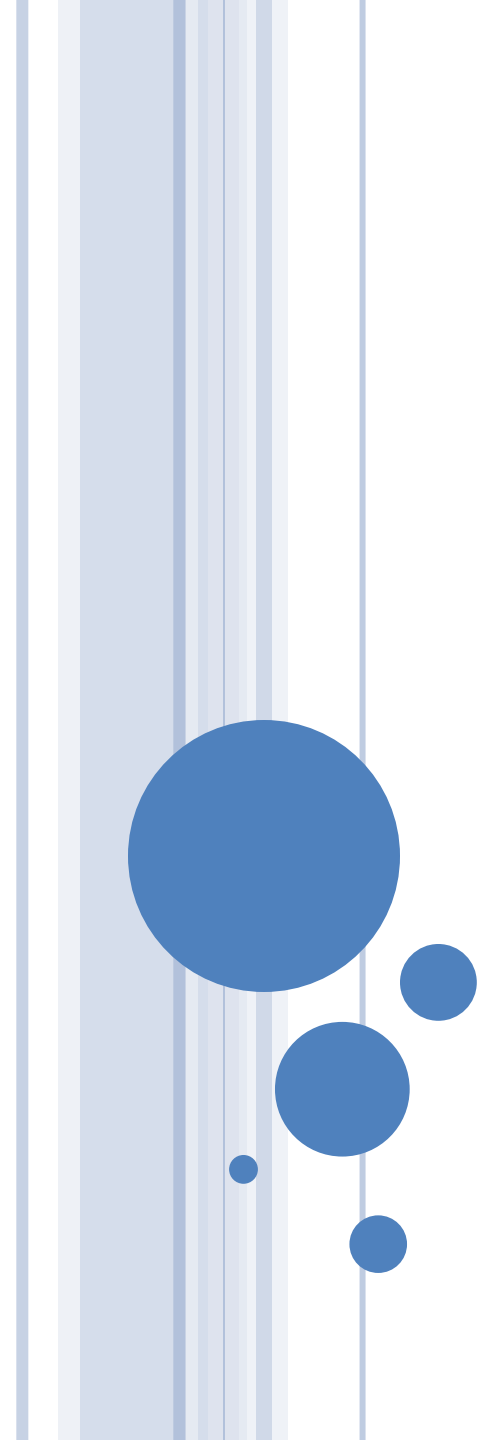




READING DIGITS IN NATURAL IMAGES WITH UNSUPERVISED FEATURE LEARNING

Fernanda Maria

Sirlene

READING DIGITS IN NATURAL IMAGES WITH UNSUPERVISED FEATURE LEARNING

- NIPS Workshop on Deep Learning and Unsupervised Feature Learning
- Ano de publicação: 2011
- Citado por 45 trabalhos.

SUMARIO

- Introdução
- SHVN Dataset
- Modelos
 - Hand Crafted Features
 - Learned Features
- Resultados
- Aplicação
 - Mapeamento
- Segmentação dos Dígitos
- Conclusão
- Referências

INTRODUÇÃO

- Processar Imagens Naturais

- Problema:

- Identificar números das casas postados nas frentes de edifícios.

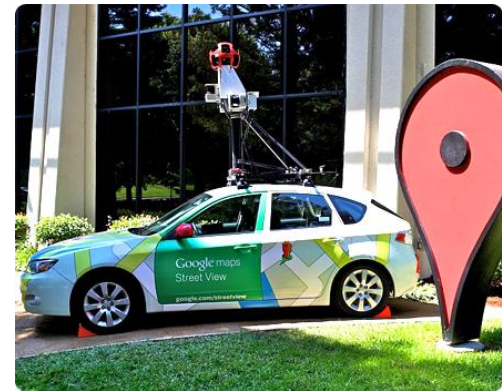


- Aplicações:

- Mapas mais precisos
- Serviços de navegação melhorados

INTRODUÇÃO

- Base:
 - Google Street View imagens (SVHN)
 - Mais de 600 mil
- Problema:
 - Dígitos de números de casas nas ruas
- Método
 - Dois Sistemas de aprendizagem
 - Grande corpus de dados de dígitos.



BASE DE DADOS: STREET VIEW HOUSE NUMBERS (SVHN)

- Mundo real (cenários naturais);
- Mais de 600.000 imagens de dígitos:
 - Treinamento: 73.257
 - Teste: 26.032
 - Extra (ligeiramente mais fáceis): 531.131
- Formatos 1:
 - Imagens originais com caixas delimitadoras.
- Formato 2 (cropped):
 - Dimensão: 32-por-32
 - Não introduz distorções de aspect ratio
 - Introduz dígitos (ou parte) além do dígito de interesse



FORMATO DOS DADOS

Formato 1

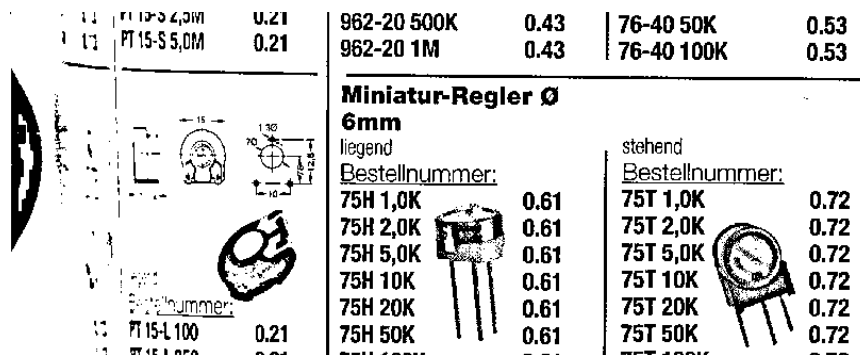


Formato 2



MODELO: HAND CRAFTED FEATURES

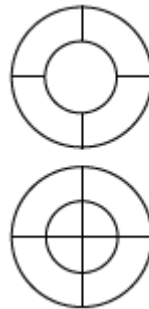
- Binarização Sauvola
 - Calcula um limiar local para cada pixel individualmente.
 - Intensidades de imagem na vizinhança local do pixel único.



MODELO: HAND CRAFTED FEATURES

○ HOG

- Captura borda ou gradientes
- Duas classes de geometrias de blocos:
 - R-HOG
 - Os quadrados ou retangulares
 - Redes densas em uma única escala sem alinhamento
 - C-HOG
 - Circulares
 - Grade de log-polar



ALGORITMOS DE APRENDIZADO DE CARACTERÍSTICAS

- Aprendem características a partir dos próprios dados;
- Características mais específicas.
- Envolvem fase não supervisionada.
- Algoritmos utilizados:
 - Auto-codificadores esparsos empilhados;
 - Rede neural convolucional (com k-means).

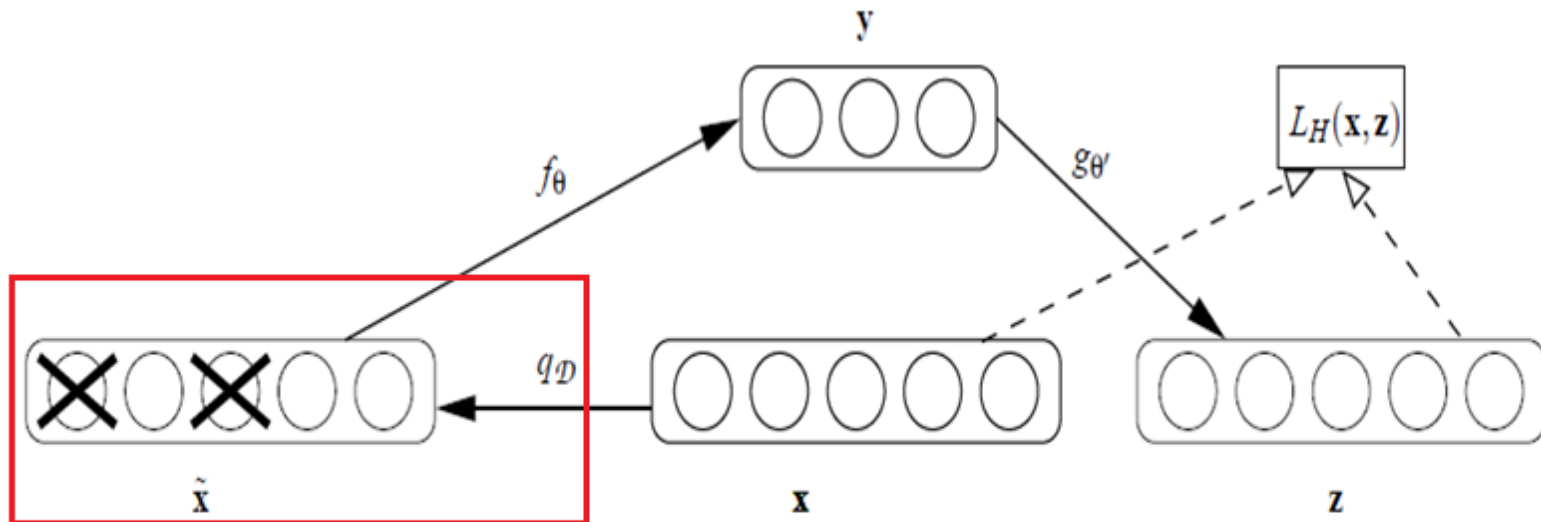


AUTO-ENCODERS ESPARSOS EMPILHADOS

- Rede não supervisionada;
- Aprende um mapeamento de suas entradas para elas próprias.
- Exemplo:
 - 8 neurônios de entrada e 3 na camada oculta
 - Auto-encoder aprende a representar 8 exemplos com codificação binária de 3 bits.



DENOISING AUTO-ENCODER

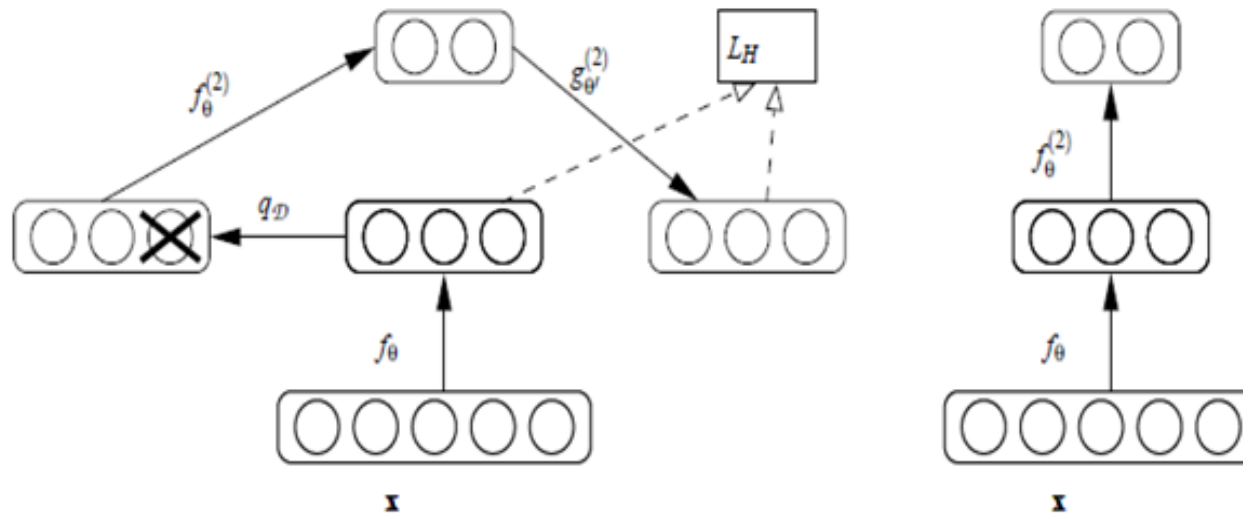


TREINAMENTO DO AUTO-ENCODER

- Cada auto-encoder é treinado usando o algoritmo de retropropagação:
 - Com o erro calculado, o algoritmo corrige os pesos em todas as camadas, partindo da saída até a entrada.
- Auto-encoder esparso: um termo para penalizar a ativação de neurônios é incluído.
- Após essa etapa de treinamento a camada de decodificação é descartada.

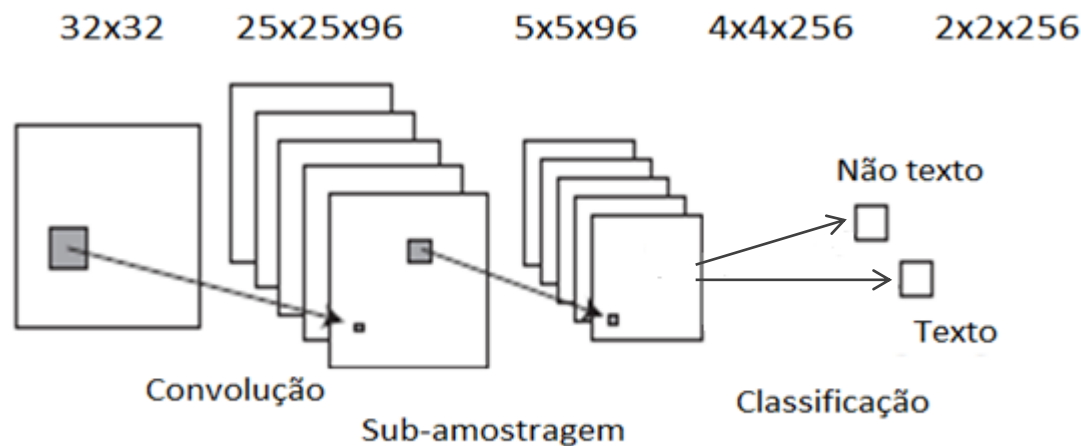


EMPILHAMENTO DE AUTO-CODIFICADORES



REDE NEURAL CONVOLUCIONAL (COM K-MEANS)

- K-means: banco de filtros D.
- Calcula-se z para cada sub-janela 8×8 .
- $z = \max \{0, |D \cdot x| - \alpha\}$, onde $\alpha = 0,5$.
- Algoritmo de retropropagação não é utilizado.
- Classificação: SVM.



RESULTADOS

ALGORITHM	SVHN-TEST (ACCURACY)
HOG	85.0%
BINARY FEATURES (WDCH)	63.3%
K-MEANS	90.6%
STACKED SPARSE AUTO-ENCODERS	89.7%
HUMAN PERFORMANCE	98.0%

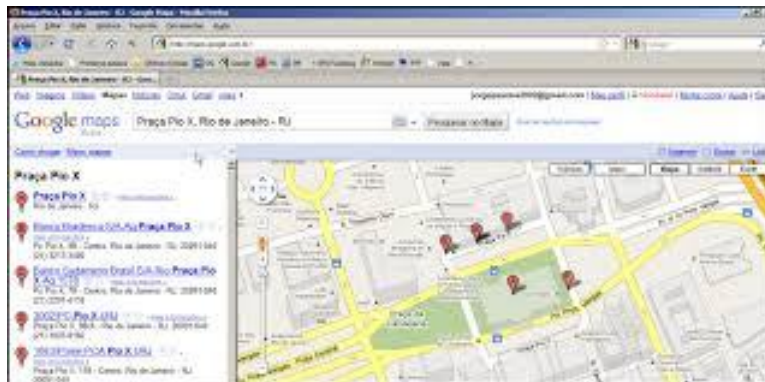
Tabela 1: precisões sobre SVHN-de teste.

APLICAÇÃO: MAPEAMENTO

- Imagens Street View
- Números de casas
 - Melhorar a precisão dos mapas.
 - Geocodificação de endereços.

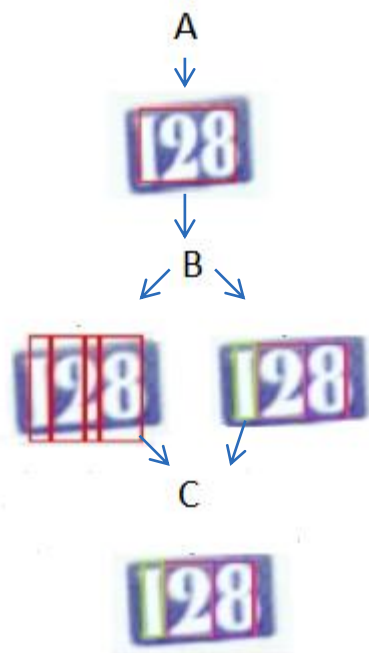


07/10/2014



SEGMENTAÇÃO DOS DÍGITOS

- A. Classificador baseado no deslizamento de janelas: intensidades da imagem e gradientes;
- B. Search Beam: usa uma função de pontuação heurística para limitar o espaço de buscas.
- C. Seleção manual.



CONCLUSÃO

- Melhores resultados: métodos que utilizam aprendizado de características não supervisionado.
- Ainda há espaço para melhorias.



REFERÊNCIAS

- Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading digits in natural images with unsupervised feature learning. In *NIPS workshop on deep learning and unsupervised feature learning* (Vol. 2011, p. 4).
- 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.
- 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.