Reconhecimento de Caracteres em Imagens com Ruído

Fernanda Maria

Sirlene Pio

Sumario

- Introdução
- Trabalhos Relacionados
- Metodologia
- Experimentos
- Conclusão
- Referências

INTRODUÇÃO

- Reconhecimento de Padrões em Imagens
  - Reconhecimento de Caracteres
    - Imagens Naturais
    - Captchas

- Problema
  - Rede Neural Convolucional
  - Duas Arquiteturas
  - Duas bases de dados

- Utilização de uma rede neural convolucional para extrair características e classificar imagens de caracteres com ruído.

Trabalhos relacionados

- Métodos de aprendizado profundo:
  - Pixels brutos -> representações gradativamente mais abstratas;
- 2011:
  - Descritores de aprendizagem profunda x extratores de caracter´ıstica hand-designed.
  - 90, 6% na base SVHN.
- 2012:
  - Rede neural convolucional tradicional + aprendizagem multi-estágio
  - 95, 1% na base SVHN.
- 2013:
  - Integração das etapas de localização, segmentação e reconhecimento
  - 97, 84% na base SVHN.

Metodologia

- Implementação
  - Alex Krizhevsky
  - Arquitetura CUDA
    - Arquitetura de Computação de Dispositivos Unificados
  - Algoritmo Back-Propagation.

Camada de Convolução

- Invariância à translação -> técnica compartilhamento de pesos;
- Técnica de compartilhamento de pesos = operação de convolução.

Camada de Max-Pooling

- Saída da camada max-pooling = máxima ativação de regiões retangulares sem sobreposição.
- Reduz o mapa de características.
- Objetivo -> selecionar características invariantes.

Camada localmente e totalmente conectadas e camada Softmax

- Camada localmente conectada

- Camada Totalmente Conectada
  - Classificação
  - N neurônio de saída = número de classes
    - 10 para a classe SVHN
    - 36 para a base CAPTCHA CNPJ

- Camada Softmax

$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Camada Regressão Logística

- Objetivo a ser otimizado -> a regressão logística multinomial.
- Regressão logística -> expressa o relacionamento entre as variáveis dependentes e independentes.

Base de Dados

- SVHN (Street View House Numbers).
  - 600 mil imagens digitais.
  - 10 classes,
  - 73.257 dígitos para treinamento,
  - 26.032 dígitos para testes, e
  - 531.131 amostras adicionais.

- Formatos
  - Imagem Original
  - Cropped

Base de dados: CAPTCHA CNPJ

- 12 mil CAPTCHAs
- 36 classes
- Os caracteres possuem:
  - Distorções
  - Tamanhos variados
  - Sobreposição
  - Muitas vezes estão incompletos.

- Os CAPTCHAs apresentam ruídos, como pontos, linhas e curvas, na mesma cor dos caracteres.

Configuração o dos Dados de Entrada para a Rede Convolucional

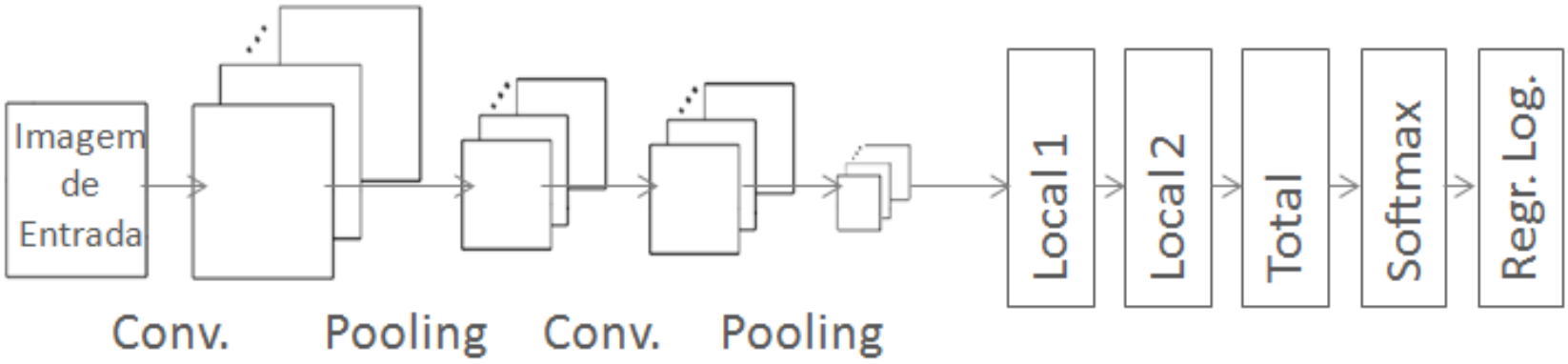
- Os dados são divididos em lotes.
- Número mínimo de lotes -> três
- Lote = matriz de dados + vetor de classes
- Cada linha da matriz de dados representa uma amostra
- O vetor de rótulos possui dimensão igual a 1-por-(número de amostras)

Experimentos

- Treinamento
  1. Lotes de treino 1 a (n-1), teste lote n.
  2. Todos os lotes de treino de 1 a n.
  3. Taxas de aprendizagem
     1. Fator de 10
     2. Treino por 10 épocas
  4. Repetição da etapa 3.

1º Arquitetura Conv-Local

- Convolução 1 + Max-pooling 1 + Convolução 2 + Max-pooling 2 + Local 1 + Local 2 + Totalmente Conectada+ Softmax + Regressão Logística.



Arquitetura Conv-Local - SVHN

SVHN			
# amostras de treinamento	# filtros	Tamanho dos filtros	Taxa de erro (%)
70k	32	5	24,95
70k	64	5	22,83
70k	128	5	22,35
70k	128	3	25,72
70k	128	7	**20,25**
150k	32	5	18,52
150k	64	5	16,49
150k	128	5	15,54
150k	128	3	18,38
150k	128	7	**14,93**
300k	32	5	14,36
300k	64	5	13,04
300k	128	5	12,91
300k	128	3	14,85
300k	128	7	**11,65**
600k	32	5	11,46
600k	64	5	10,28
600k	128	5	9,97
600k	128	3	11,45
600k	128	7	**9,35**

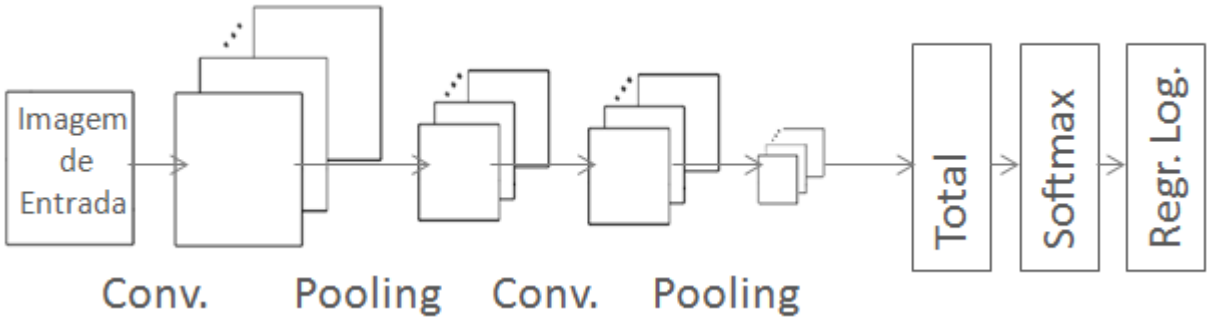
Arquitetura Conv-Local – Captcha CNPJ

- Parâmetros utilizados:
  - Aqueles que obtiveram a menor taxa de erro na base SVHN.

CAPTCHA CNPJ			
# amostras de treinamento	# filtros	Tamanho dos filtros	Taxa de erro (%)
10k	128	7	5,94

2º Arquitetura Conv

- Convolução 1 + Max-pooling 1 + Convolução 2 + Max-pooling 2 + Totalemnte Conectada+ Softmax + Regressão Logística.



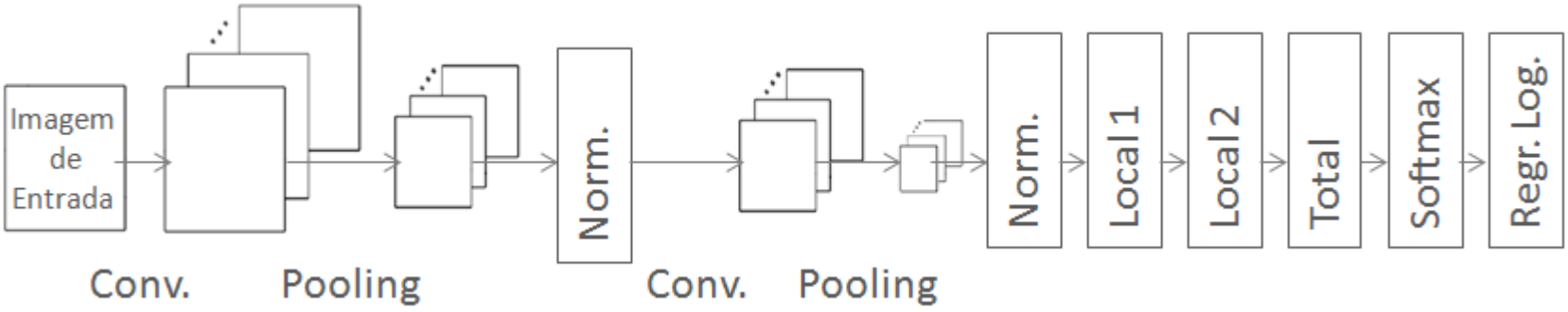
Arquitetura Conv

- Objetivo: verificar e quantificar a influência da presença ou ausência das camadas localmente conectadas na rede neural de convolução.

Base de dados	# amostras de treinamento	# filtros	Tamanho dos filtros	Taxa de erro (%)
SVHN	600k	128	7	13,49
CAPTCHA CNPJ	10k	128	7	4,58

3º Arquitetura Conv-Norm-Local

- Convolução 1 + Max-pooling 1+ Normalização1 + Convolução 2 + Max-pooling 2 + Normalização2 + Local1 + Local2+ Totalemente Conectada + Softmax + Regressão Logística.



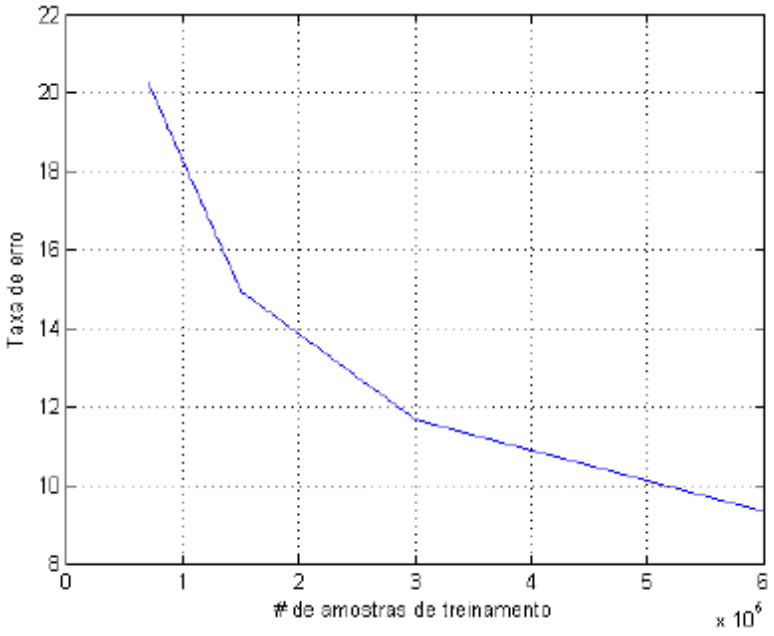
3º Arquitetura Conv-Norm-Local

- Resultados
  - Base de dados CAPTCHA CNPJ
  - Aumento gradativo do número de filtros da camada de convolução:
    - objetivo: diminuir a taxa de erro de acordo com oss primeiros resultados.

CAPTCHA CNPJ			
# amostras de treinamento	# filtros	Tamanho dos filtros	Taxa de erro (%)
10k	64	7	4,42
10k	128	7	**4,07**
10k	192	7	4,33

Análise dos Resultados

- Base SVHN:
  - Menor taxa de erro:
    - 9, 35%
    - 128 filtros de dimensões 7-por-7
  - Maior taxa de erro:
    - 25,72%
    - 128 filtros de dimensões 3-por-3
- Base Capcha
  - Conv-Norm-Local,
  - Menor taxa de erro:
    - 4; 07%
    - 128 filtros de dimensões 7-por-7



Conclusão

- Diferentemente dos resultados obtidos na base SVHN, na base CAPTCHA CNPJ, a arquitetura sem camadas localmente conectadas obteve uma taxa de erro menor que a arquitetura com tais camadas.
- Baseando-se na redução gradativa observada na taxa de erro `a medida que o conjunto de treinamento cresce, será realizado um aumento do número de amostras de treinamento gerando imagens sintéticas, através da aplicação de rotações, espelhamento, inserção de ruídos, nas imagens originais.

Referências

1. X. S. Canto, F. M. Ramirez, and V. U. Cetina. Parallel training of a back-propagation neural network using cuda. In Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on, pages 307–312. IEEE, 2010.

2. A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T.Wang, D. J.Wu, and A. Y. NG. Text detection and character recognition in scene images with unsupervised feature learning. In Document Analysis and Recognition (ICDAR), 2011 International Conference on, pages 440–445. IEEE, 2011.

3. Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading digits in natural images with unsupervised feature learning. In NIPS workshop on deep learning and unsupervised feature learning, volume 2011, page 4, 2011.