Aprendizado profundo com capacidade computacional reduzida: uma aplicação à quebra de CAPTCHAs.

Diogo Felipe Félix de Melo

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

Identificação

Aluno(a): Diogo Felipe Félix de Melo.

Orientador(a): Pablo de Azevedo Sampaio.

Título: Aprendizado profundo com capacidade computacional reduzida: uma aplicação à quebra de captchas.

Área de Concentração: Aprendizado de Máquina.

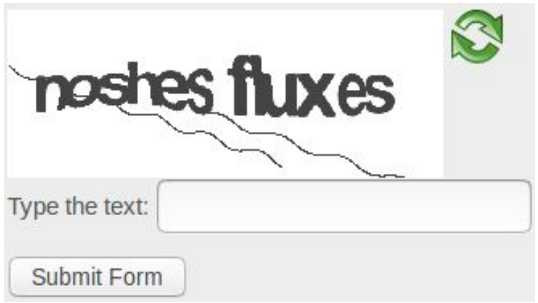
Linha de Pesquisa: Redes Neurais de aprendizado profundo.

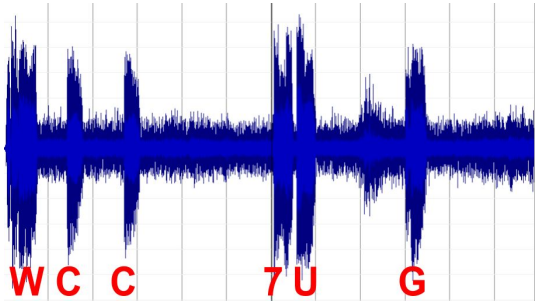
Agenda

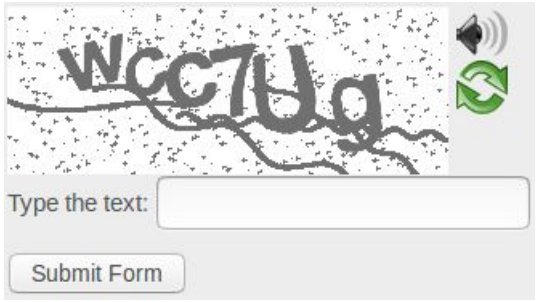
- CAPTCHAs
- Resultados na literatura
- Método Proposto
- Resultados
- Conclusões

CAPTCHAs

CAPTCHAS







Diferentes tipos de captcha gerados com a biblioteca Securimage.
Créditos: O autor.

CAPTCHAs de Texto

Captcha de texto gerado com a biblioteca SimpleCaptcha. Créditos: o autor

CAPTCHAs de Texto

Captcha de texto gerado com a biblioteca SimpleCaptcha. Créditos: o autor

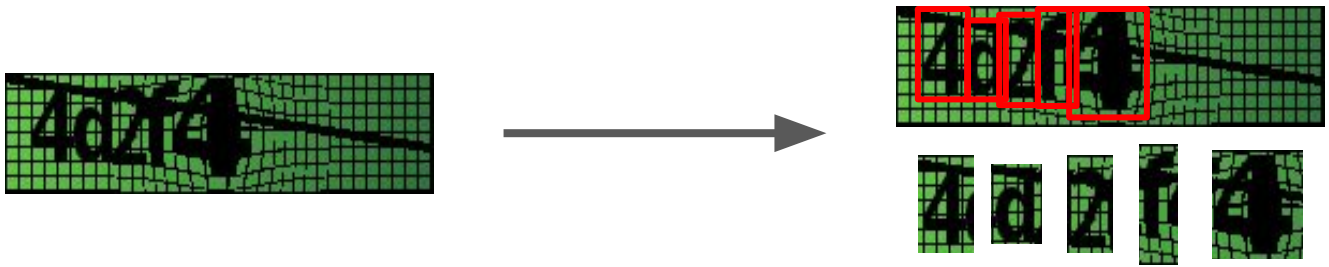
Resultados na literatura

Resultados na literatura

- Informada: Conhecimento prévio do problema, projeto de pipeline para tratamento de imagem.

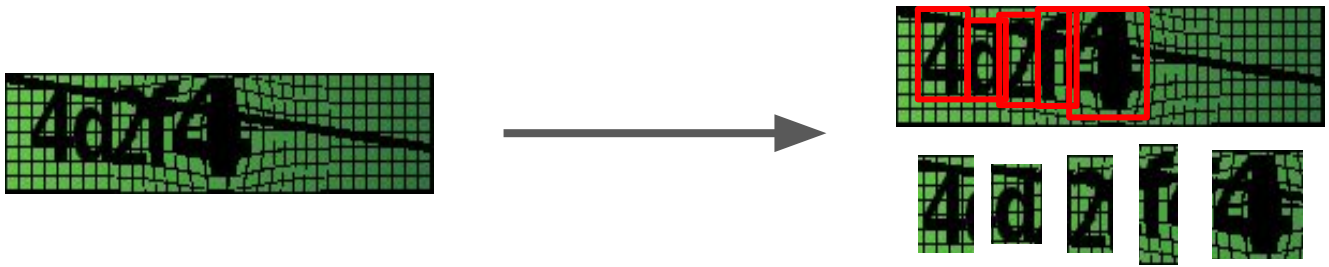
Resultados na literatura

- Informada: Conhecimento prévio do problema, projeto de pipeline para tratamento de imagem.



Resultados na literatura

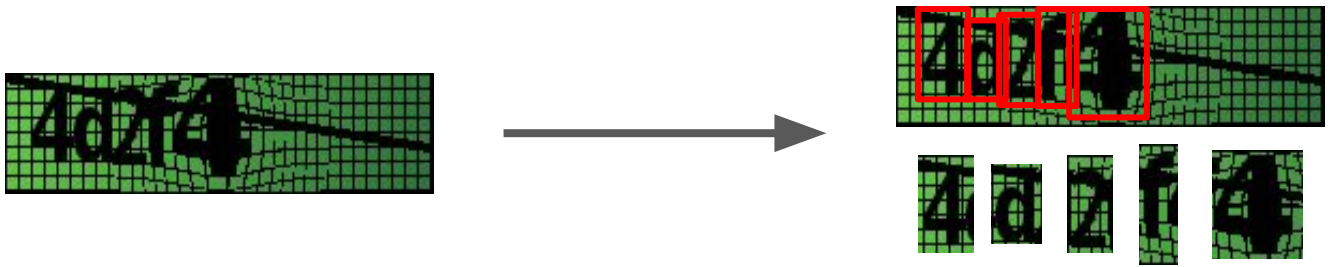
- Informada: Conhecimento prévio do problema, projeto de pipeline para tratamento de imagem.



- Não informada: Mapa imagem -> texto visto como operador único.

Resultados na literatura

- Informada: Conhecimento prévio do problema, projeto de pipeline para tratamento de imagem.



- Não informada: Mapa imagem -> texto visto como operador único.

Resultados na literatura

Referência	Resultados	Limitação
Goodfellow et al., 2013	96% no SVHN (NETZER, Y. et al., 2011), 99.8% no reCaptcha v2	De 9 a 11 camadas convolucionais (8-192 filtros), camada densa com mais de 3 mil entradas, mais de 500 mil exemplos.
Pinto, 2016	77% de acerto em base sintética.	180 mil exemplos, quatro camadas convolucionais com 64, 128, 256 e 512 canais, duas camadas densas com mais de 4000 entradas, totalizando mais de 60 milhões parâmetros.
George et al, 2017	Boas taxas de acertos com poucos exemplos.	Treinamento de difícil paralelização. Muitas horas de treino. Necessidade de acompanhamento por humanos.

Resultados na literatura

Referência	Resultados	Limitação
Goodfellow et al., 2013	96% no SVHN (NETZER, Y. et al., 2011), 99.8% no reCaptcha v2	De 9 a 11 camadas convolucionais (8-192 filtros), camada densa com mais de 3 mil entradas, mais de 500 mil exemplos.
Pinto, 2016	77% de acerto em base sintética.	180 mil exemplos, quatro camadas convolucionais com 64, 128, 256 e 512 canais, duas camadas densas com mais de 4000 entradas, totalizando mais de 60 milhões parâmetros.
George et al, 2017	Boas taxas de acertos com poucos exemplos.	Treinamento de difícil paralelização. Muitas horas de treino. Necessidade de acompanhamento por humanos.

Abordagem proposta

Abordagem proposta (princípios)

1. O início da dinâmica de uma configuração fornece informações importantes sobre o comportamento seu ao longo do resto treino.

2. Configurações similares tendem a produzir resultados similares.

3. Uma experimentação mais consistente resulta em conclusões mais consistentes.

Abordagem proposta (método)

1. Várias configurações são treinadas durante um tempo reduzido e sua performance avaliada.

2. As arquiteturas à serem estudadas são substancialmente diferentes entre sí. Esse princípio também é aplicado à parâmetros contínuos como a taxa de aprendizado, por exemplo.

3. Um conjunto de hiperparâmetros sempre será mantido fixo enquanto os demais são explorados de forma mais minuciosa, permitindo uma comparação direta do impacto de cada escolha.

Abordagem proposta (experimentos)

Arquiteturas: M[D], C6M[D][Max], C6C12M[D][Max], C6C12Fl100M[D][Max], C6C12C36C36M[D][Max], C6C12C36C36Fl100 M[D][Max].

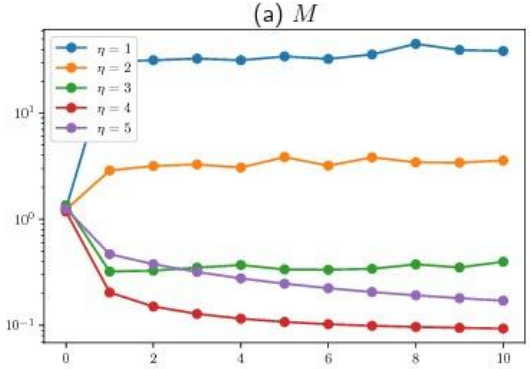
Taxa de aprendizado: $10^{-\eta}$, com $\eta = 1, 2, 3, 4, 5$.

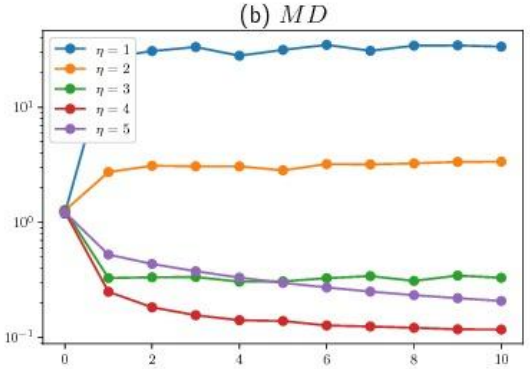
Configurações treinadas durante 10 épocas.

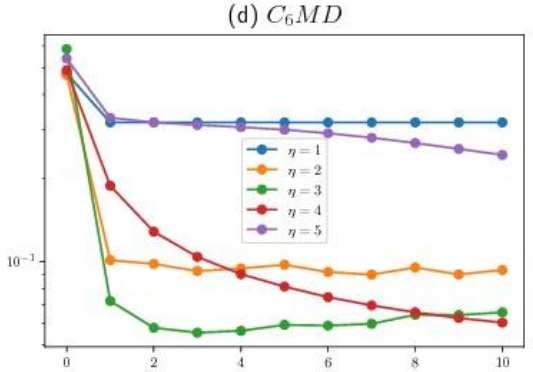
Configuração escolhida treinada por mais épocas.

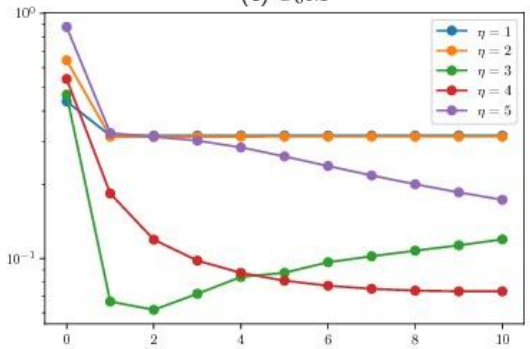
Resultados

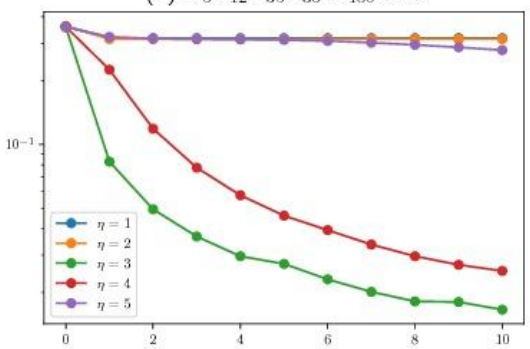
Resultados (experimentos iniciais)

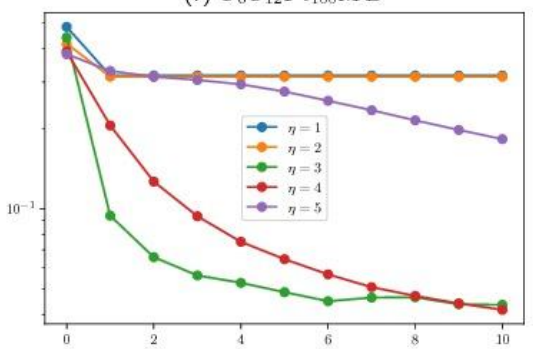




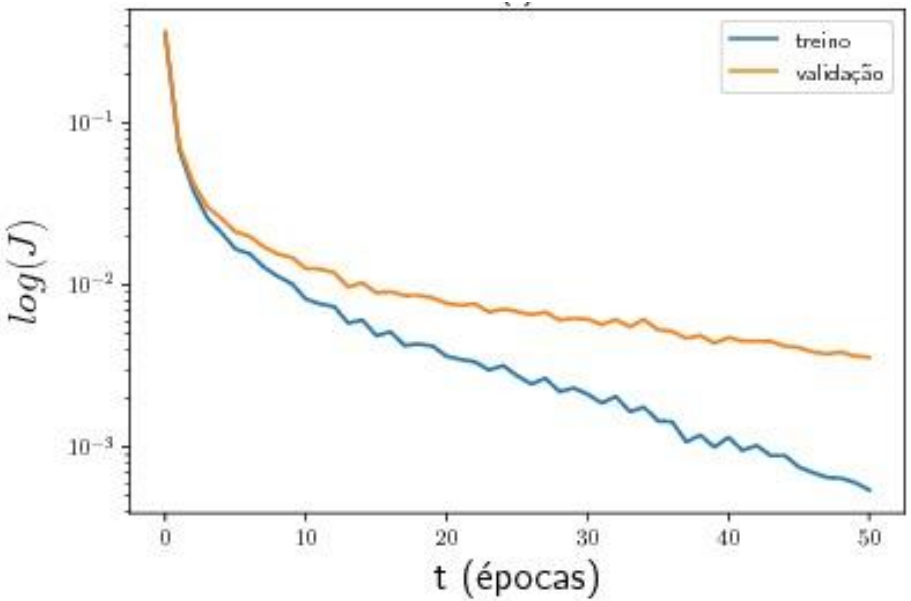


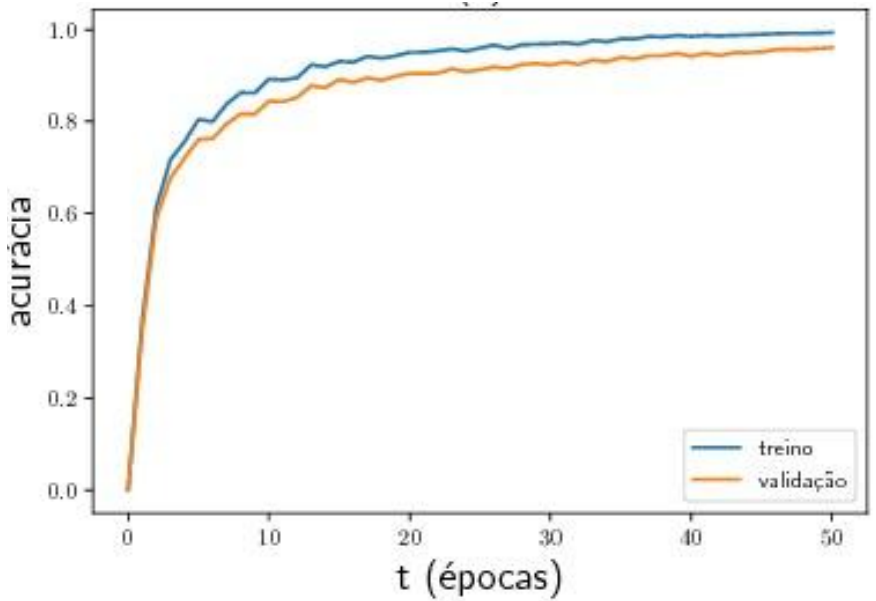






Resultados (modelo C6C12C36C36FI100MD)





Conclusões

Conclusões

1. A abordagem proposta se mostrou efetiva para escolha dos hiperparâmetros.

2. Por projeto, pode ser aplicada de forma iterativa e respeita as restrições do ambiente.

3. Modelo final com precisão de 96% e 290 mil parâmetros, treinado em pouco mais de 3 horas.

Obrigado!

Referências

GOODFELLOW, I. J. et al. Multi-digit number recognition from street view imagery using deep convolutional neural networks. CoRR, abs/1312.6082, 2013. Disponível em:<http://arxiv.org/abs/1312.6082>

GEORGE, D. et al. A generative vision model that trains with high data efficiency and breaks text-based captchas. Science, American Association for the Advancement of Science, 2017. ISSN 0036-8075. Disponível em: <http://science.sciencemag.org/content/early/2017/10/26/science.aag2612>

PINTO, V. A. Redes Neurais Convolucionais de Profundidade Para Reconhecimento de Texto em Imagens de CAPTCHA. Florianópolis, Santa Catarina.: UNIVERSIDADE FEDERAL DE SANTA CATARINA - DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA, 2016. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/171436>

NETZER, Y. et al. Reading digits in natural images with unsupervised feature learning. In: NIPS workshop on deep learning and unsupervised feature learning. [S.l.: s.n.], 2011. v. 2011, n. 2, p. 5.