

Ademir Rafael Marques Guedes
Victor Guimaraes

Convolutional Neural Networks
Applied to
House Numbers Digit Classification

Pierre Sermanet

Soumith Chintala

Yann LeCun

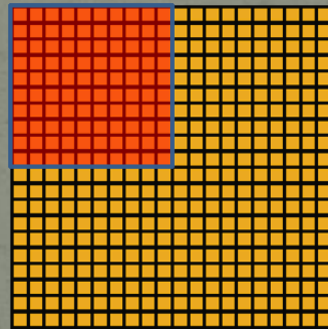
The Courant Institute of Mathematical Sciences - New York University

The Street View House Numbers (SVHN) Dataset

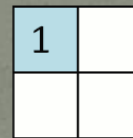


Pooling

- Técnica de sumarização de um conjunto de características em um conjunto bem menor, reduzindo assim o volume de dados(características) a ser processado



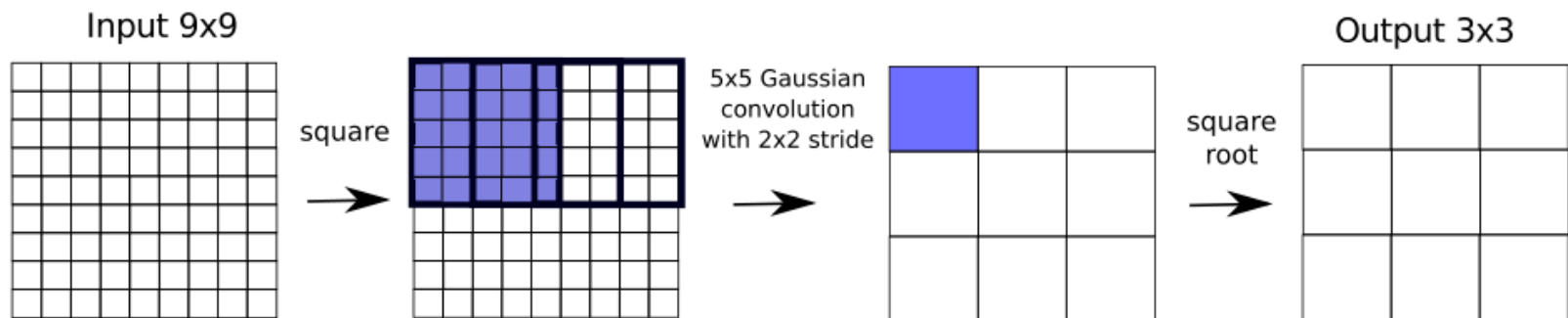
Convolved
feature



Pooled
feature

LP-Pooling

- Método de sumarização de características inspirado em células complexas do aparelho visual (inspirado nas características biológicas da visão)
- Dá maior importância a características mais fortes e suprime características mais fracas da imagem



LP-Pooling

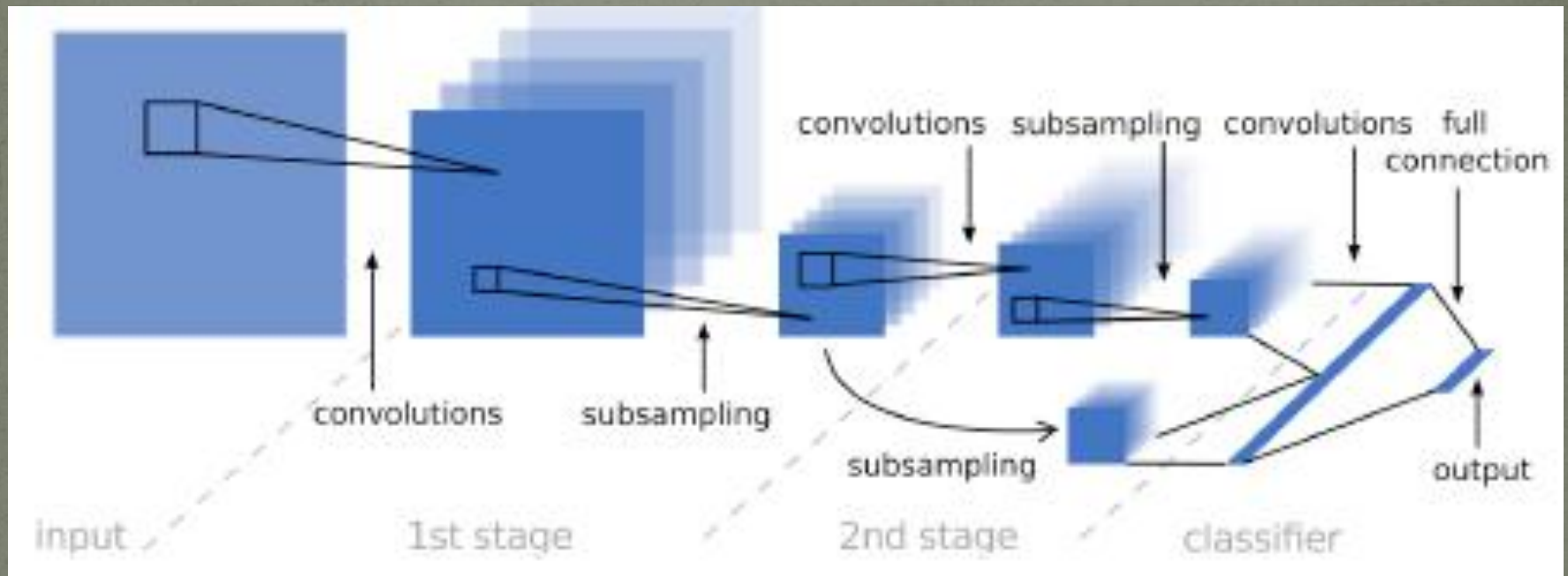
$$O = (\sum \sum I(i, j)^P \times G(i, j))^{1/P}$$

- O – matriz de saída
- I – matriz de entrada
- G – kernel Gaussiano de I
- P – parâmetro de entrada a ser definido

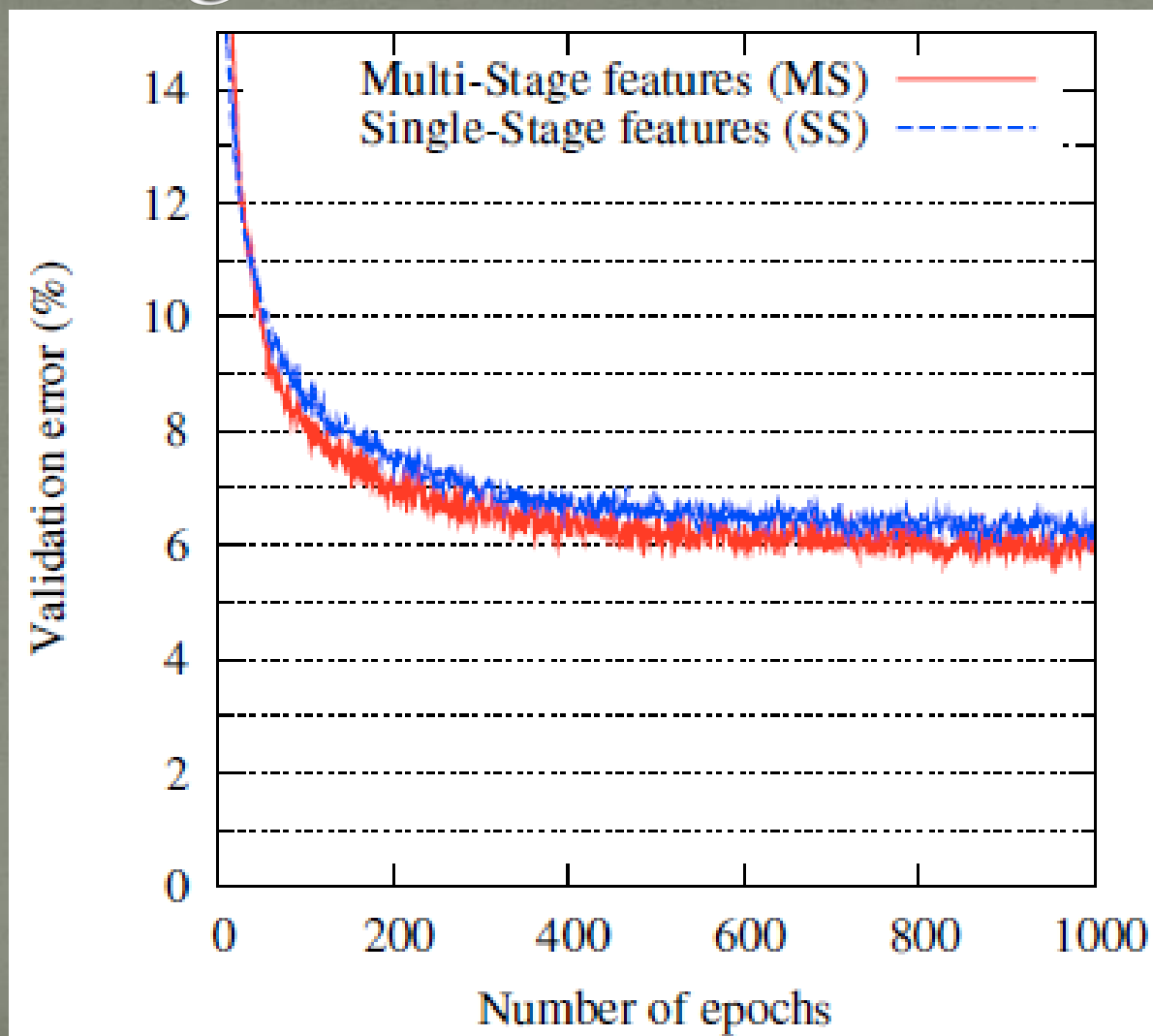
Multi-stage feature

- Geram informações mais ricas contendo:
 - Texturas locais;
 - Detalhes finos perdidos por níveis mais elevados;
- Estágios compostos por uma convolução e uma sub-amostragem;
- Ramifica saídas de todos os estágios para o classificador;

Multi-stage feature



Multi-stage feature



Multi-stage feature

Task	Single-Stage features	Multi-Stage features	Improvement %
Pedestrians detection (INRIA) [10]	14.26%	9.85%	31%
Traffic Signs classification (GTSRB) [12]	1.80%	0.83%	54%
House Numbers classification (SVHN)	5.54%	5.36%	3.2%

Experimentos

- Imagens de 32x32 com 3 canais;
- Imagens divididas em 3 conjuntos:
 - Treino: menor e difícil;
 - Extra: maior e fácil;
 - Teste;

Experimentos

- Conjunto de validação:
 - ordem aleatória;
 - $\frac{2}{3}$ vindos do conjunto de treinamento;
 - $\frac{1}{3}$ do conjunto extra;
 - Total: 6000 elementos;
- Conjunto de treino: 598388 elementos;
- Conjunto de teste: 26032 elementos;

Experimentos

- Normalização de contraste local
 - Kernel 7x7;
 - Aplicado somente no primeiro canal;
- Normalização de contraste global:
 - Sobre cada canal;
- Amostras sem distorção;
- Moldura de 2 pixels em alguns experimentos;

Experimentos

- A rede de convolução tem 2 estágios de extração de características e um classificador não-linear de duas camadas
- Primeira camada de convolução produz 16 características
- Segunda camada produz 512 características

Experimentos

- Foram realizados testes com $P = 1, 2, 4, 8, 12, 16, 32$ e ∞ no conjunto de validação e o valor de P com melhor performance é utilizado no conjunto de testes
- Para a base utilizada, $P = 2, 4$ e 12 obtiveram melhores resultados

Experimentos

Algorithm	SVHN-Test Accuracy
Binary Features (WDCH)	63.3%
HOG	85.0%
Stacked Sparse Auto-Encoders	89.7 %
K-Means	90.6%
ConvNet / MS / Average Pooling	90.94%
ConvNet / MS / L2 / Smaller training	91.55%
ConvNet / SS / L2	94.46%
ConvNet / MS / L2	94.64%
ConvNet / MS / L12	94.89%
ConvNet / MS / L4	94.97%
ConvNet / MS / L4 / Padded	95.10%
Human Performance	98.0%