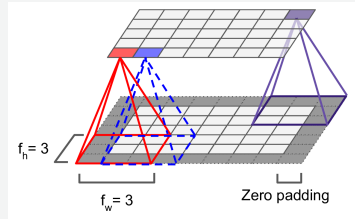


2017-12-17

Systemy agentowe

└ Filtr/jądro (ang. *filter/kernel*)

Filtr/jądro (ang. *filter/kernel*)



A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow* 2017

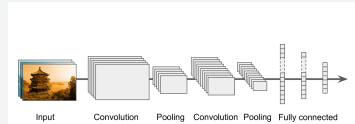
Filtr to po prostu suma ważona pikseli pod spodem (+bias), bez nieliniowości

2017-12-17

Systemy agentowe

└ Typowa architektura

Typowa architektura

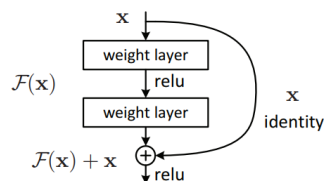


A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow* 2017

Po convolution zwykle jest nieliniowość, np. ReLU

Residual learning

Residual learning



K. He et al., *Deep Residual Learning for Image Recognition*
<https://arxiv.org/pdf/1512.03385v1.pdf>

$F(x)$ uczy się tylko różnicy pomiędzy x a oczekiwanym wyjściem, co jest łatwiejsze do optymalizacji

Architektura sieci

Architektura sieci

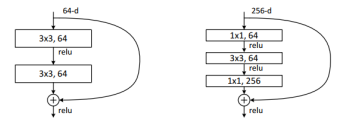
layer name	output size	104-layer	54-layer	38-layer	104-layer	152-layer	
conv1	112x112	3x3 conv, stride 2, padding 1					
conv2A	56x56	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		12	12	12	12	12	
conv2B	56x56	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		12	12	12	12	12	
conv3A	28x28	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		16	16	16	16	16	
conv3B	28x28	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		16	16	16	16	16	
conv4A	14x14	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		8	8	8	8	8	
conv4B	14x14	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		8	8	8	8	8	
conv5A	7x7	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		4	4	4	4	4	
conv5B	7x7	3x3 conv	3x3 conv	3x3 conv	3x3 conv	3x3 conv	
		4	4	4	4	4	
fc6	1000	average pool, 1000-d fc, softmax					
fc7	1000	1000-d fc, softmax					

K. He et al., *Deep Residual Learning for Image Recognition*
<https://arxiv.org/pdf/1512.03385v1.pdf>

Nawiasy kwadratowe oznaczają bloki residual learning (tzn. takie jak pokazano na poprzednim slajdzie).

└ Bottleneck architecture

Bottleneck architecture



K. He et al., *Deep Residual Learning for Image Recognition*
<https://arxiv.org/pdf/1512.03386v1.pdf>

Po lewej stronie zwykła architektura, po prawej stronie wariant z bottleneck: pierwsza warstwa zmniejsza liczbę wymiarów, a ostatnia ją odtwarza, dzięki temu środkowa warstwa (z filtrem 3×3) ma mniejsze wejście i wyjście. Po lewej $3 \cdot 3 \cdot 64 \cdot 64 \cdot 2 = 73728$ parametry, po prawej $64 \cdot 256 + 3 \cdot 3 \cdot 64 \cdot 256 + 256 \cdot 64 = 69632$ parametrów: efektywniejsze obliczeniowo, a wydajnościowo podobne.