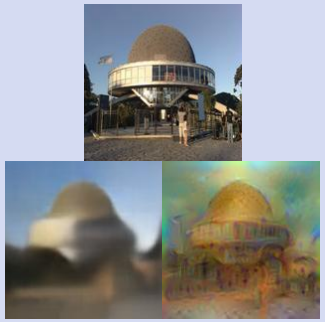
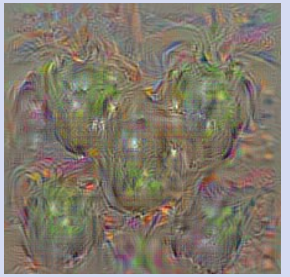
Feature map inversion



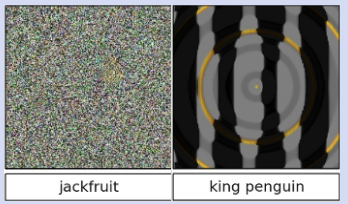
Inverting conv5 (AlexNet)
(Left - Dosovitskiy & Brox, 2015)
(Right - Mahendran & Vedaldi, 2014)

Activation optimisation

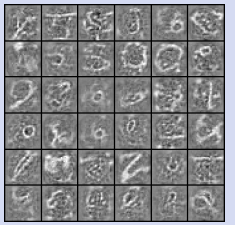
Activation optimisation using DGN as natural prior
(Nguyen et al, 2016)



Activation maximisation
(Simonyan et al, 2013)



Fooling DNNs (Nguyen et al, 2014)



Activation maximisation
(Erhan et al, 2009)

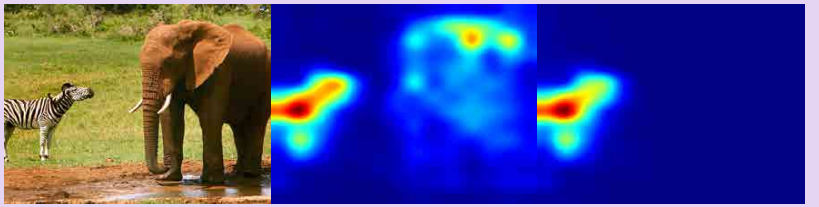
Natural prior regularisation

Inverting relu3 (AlexNet) feature maps, regularisation effects:
Left: Jitter, Right: Total variation
(Mahendran & Vedaldi, 2014)

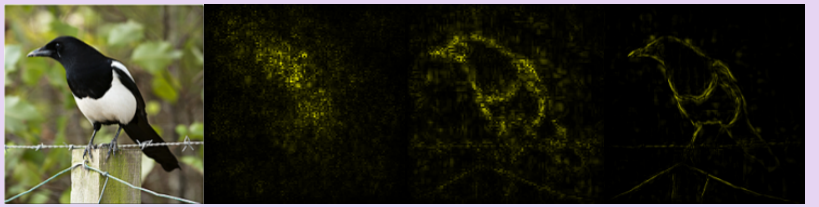
Caricaturing

Caricaturing "red fox"
at layer relu4 (VGG-M)
(Mahendran & Vedaldi, 2016)

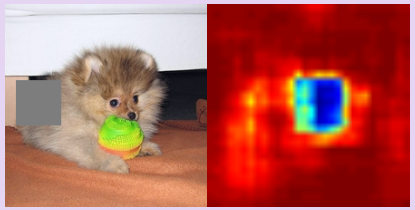
Attention mapping



Excitation backpropagation, non-contrastive (left), contrastive (right)
(Zhang et al. 2016)



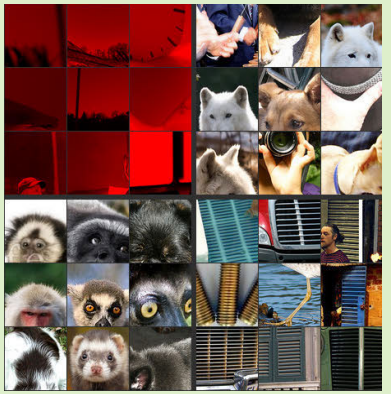
Saliency, Deconvolution, LRP attention maps
(Simonyan 2013, Zeiler & Fergus 2013, Bach et al. 2015)
(figure by Samek et al. 2015)



Occlusion study
(Zeiler & Fergus, 2013)

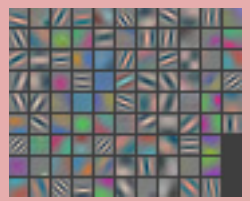
Dataset-centric

t-SNE ImageNet validation images
(visualisation by Andrej Karpathy)
(t-SNE by Maaten et al. 2008)

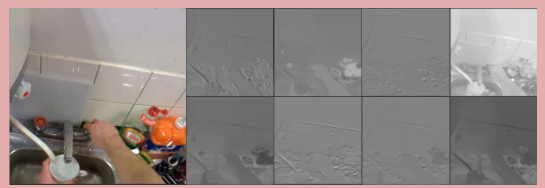


Example activation maximisation
top-9 examples for 4 neurons
(Zeiler & Fergus, 2013)

Filter Analysis



AlexNet Filter weights
(Zeiler & Fergus, 2013)



Filter responses of
VGG-16 conv1
obtained with Deep
Visualisation Toolbox
(Yosinski et al, 2014)