Visualizing and understanding convolutional neural networks

Object Recognition and Computer Vision - 2018

Éloïse Berthier & Clément Mantoux

Introduction

Deep convnets perform extremely well on images

But little explainability

→ Deconvolutional NN to visualize convnets

→ Tool for network architecture design

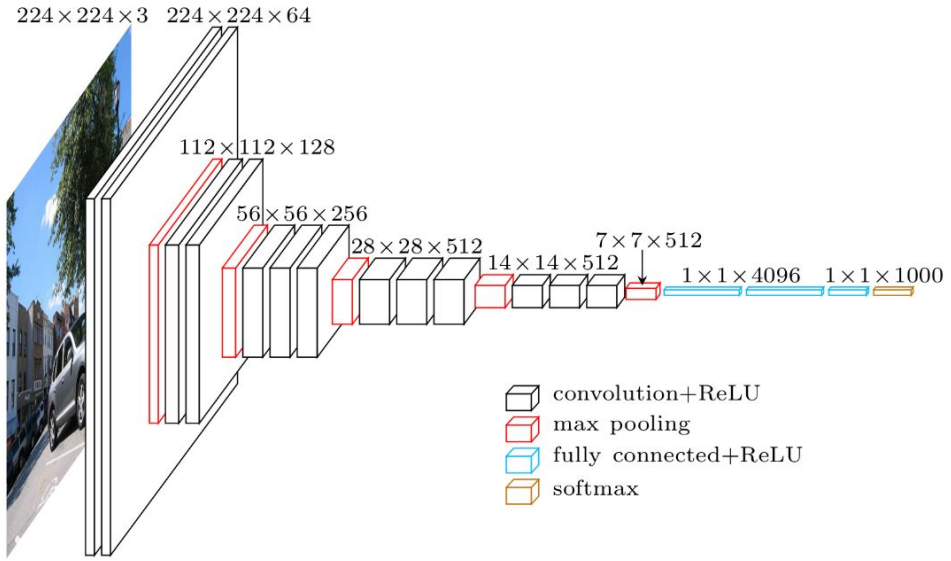

Figure : the VGG16 network architecture

Visualizing convnets with deconvnets

Deconvolutional Neural Networks:

- Convolution → Transposed Convolution
- Pooling → Unpooling with argmax
- ReLU → ReLU

Convnet visualization:
→ Nullify all channels but one and feed in the deconvnet

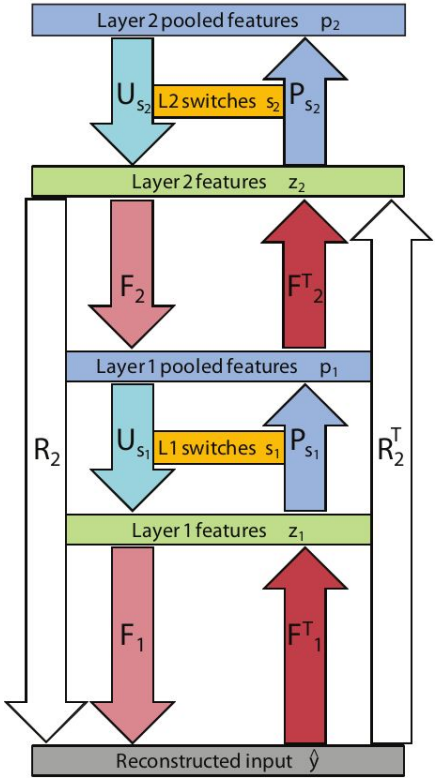

Figure : deconvnet from [Zeiler *et al.* ICCV 2011]

Features Visualization on VGG16

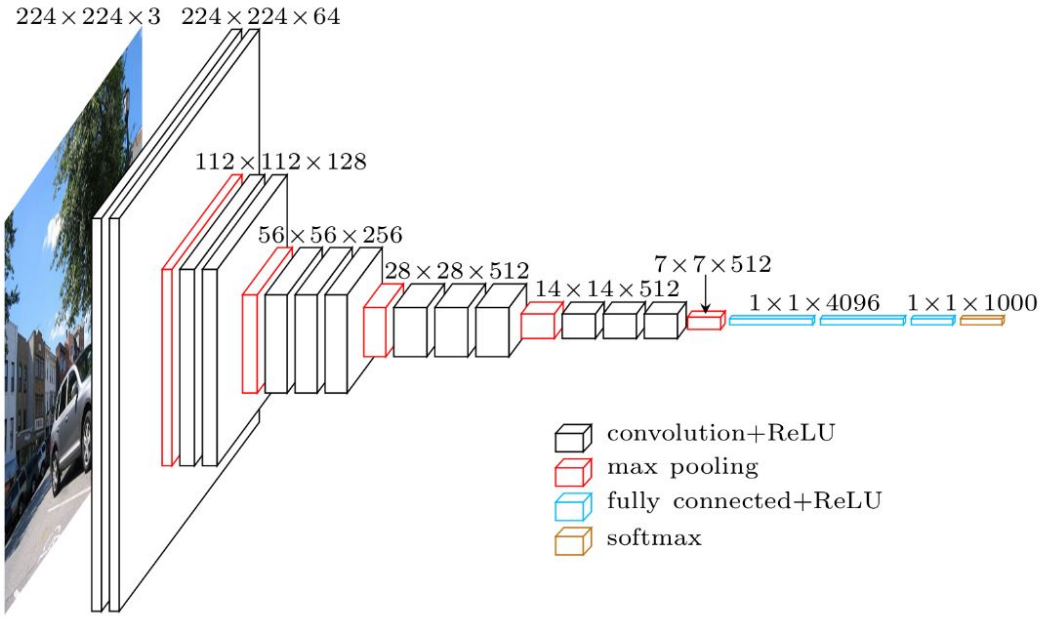

Figure : the VGG16 network architecture

Features Visualization on VGG16

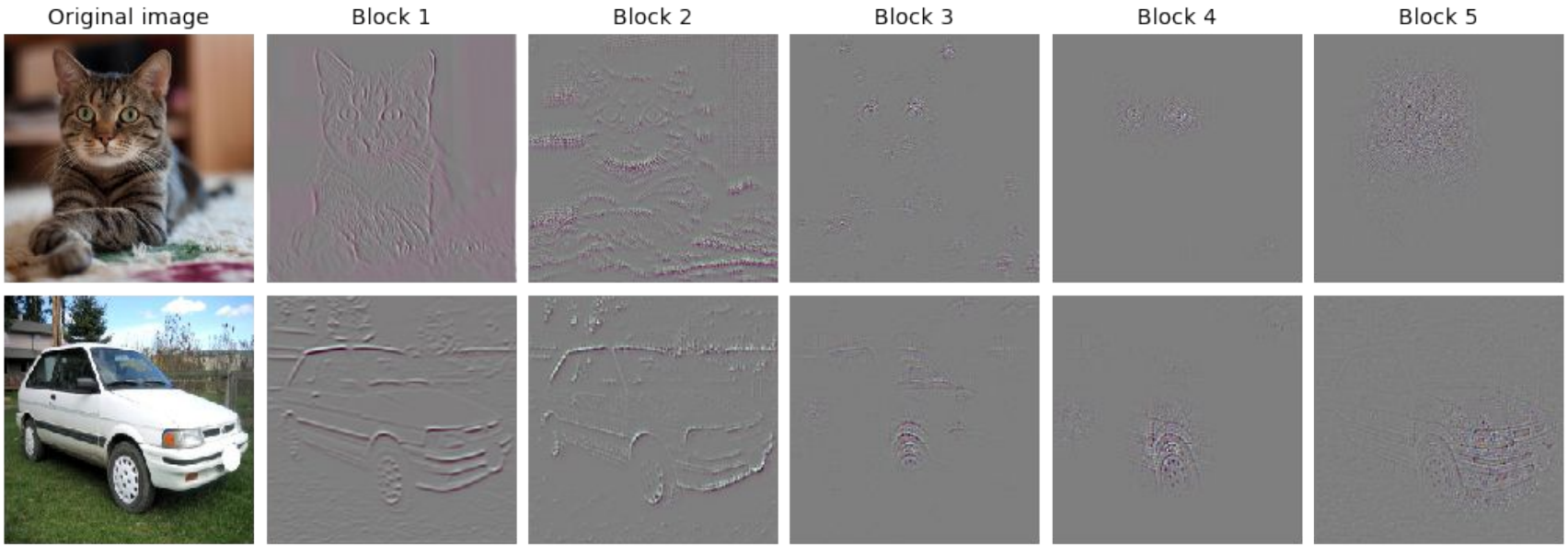

Figure : Best feature map at each network block

Training a ConvNet on CIFAR-10

32x32 images

10 classes

50k train images

10k test images

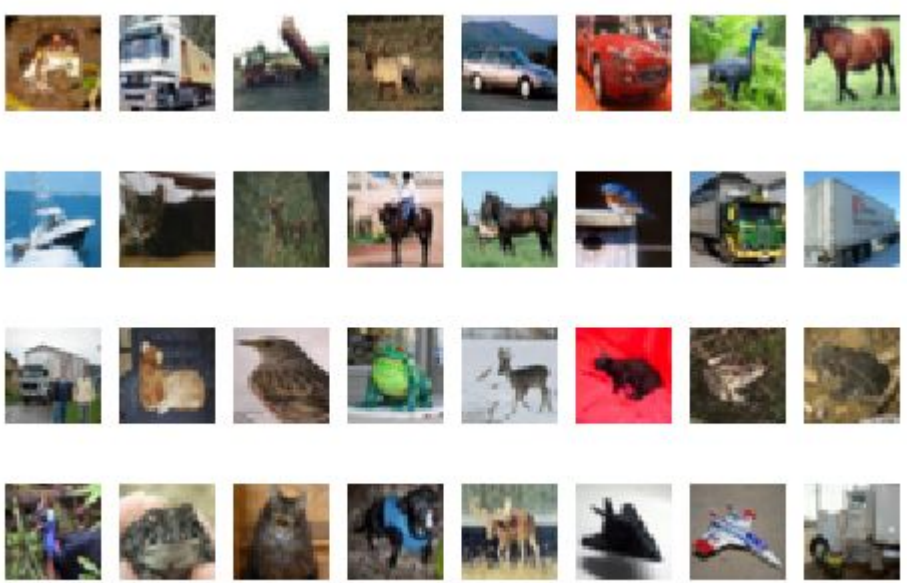

Figure: Sample images from CIFAR-10

Batch Normalization & Padding

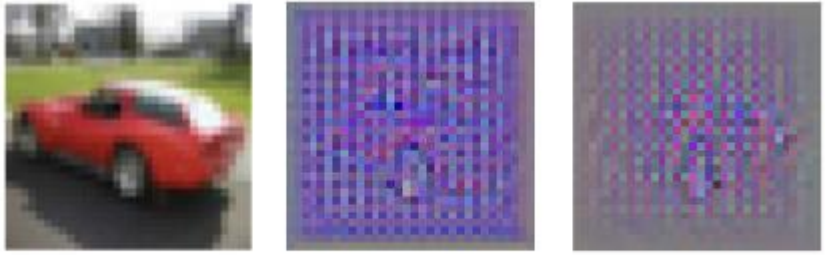

Figure: Most activated filters of layer 4

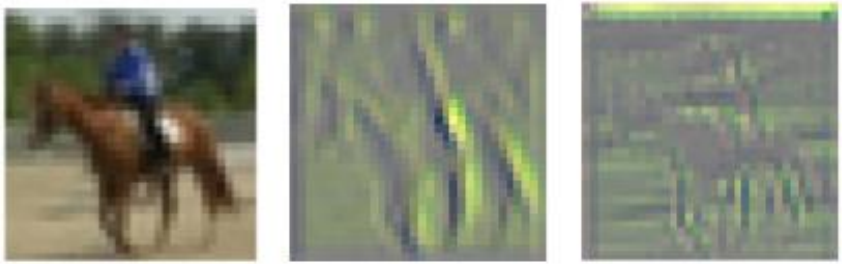

Figure: Most activated filters of layer 1

Evolution of filters during training

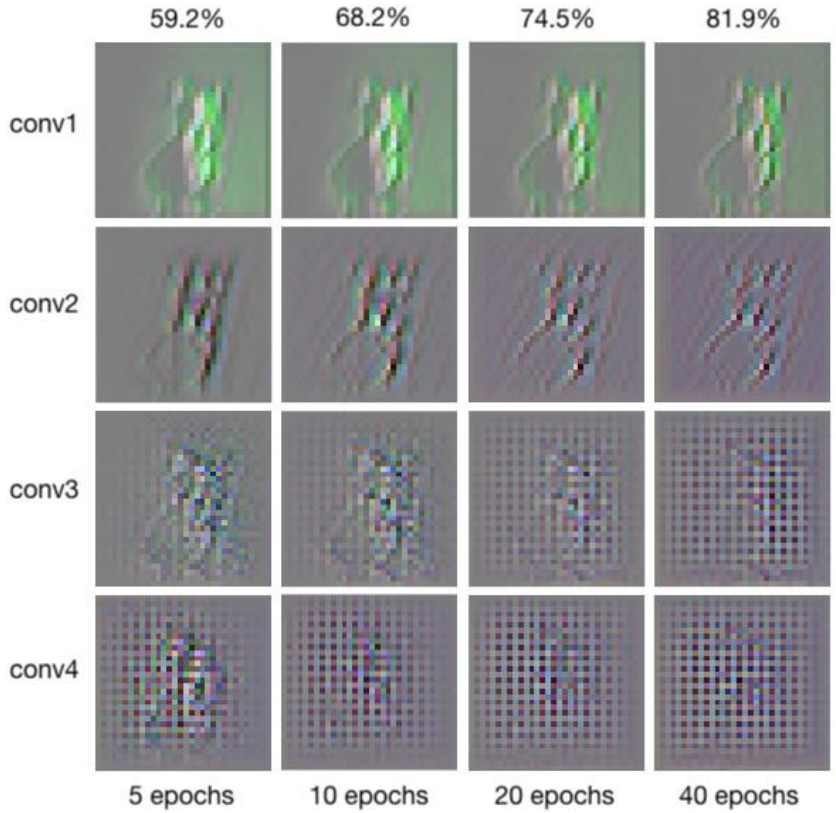

Figure: Tracking of 1 filter per layer during training

Checkerboard effect

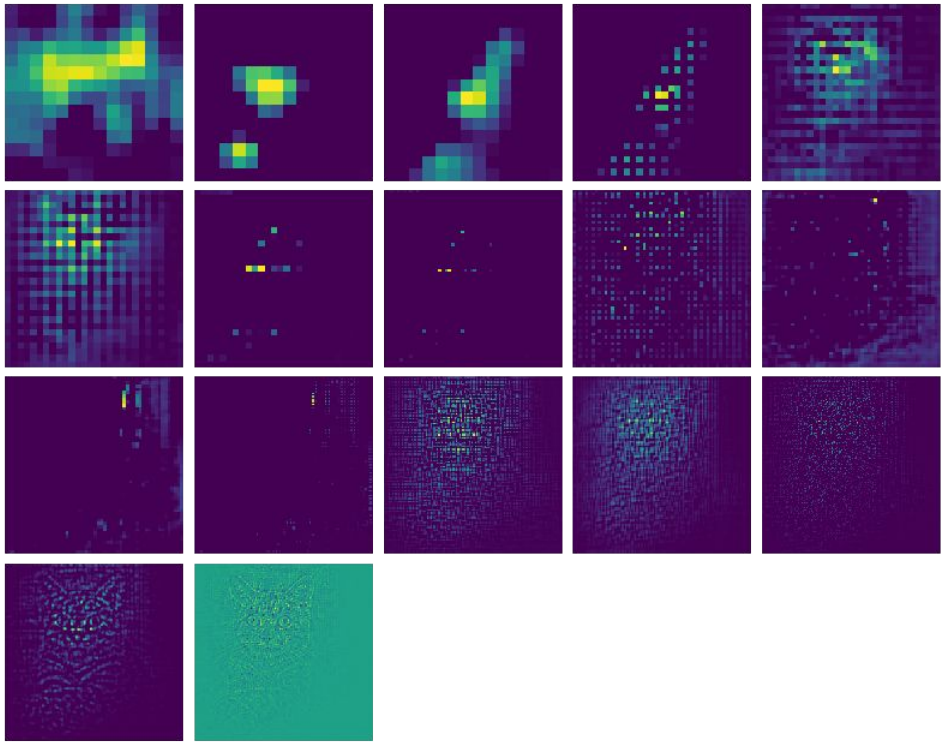

Figure: Activation of VGG16 deconvnet on each layer from the deepest one

Final architecture: 81.9% test accuracy

588k parameters

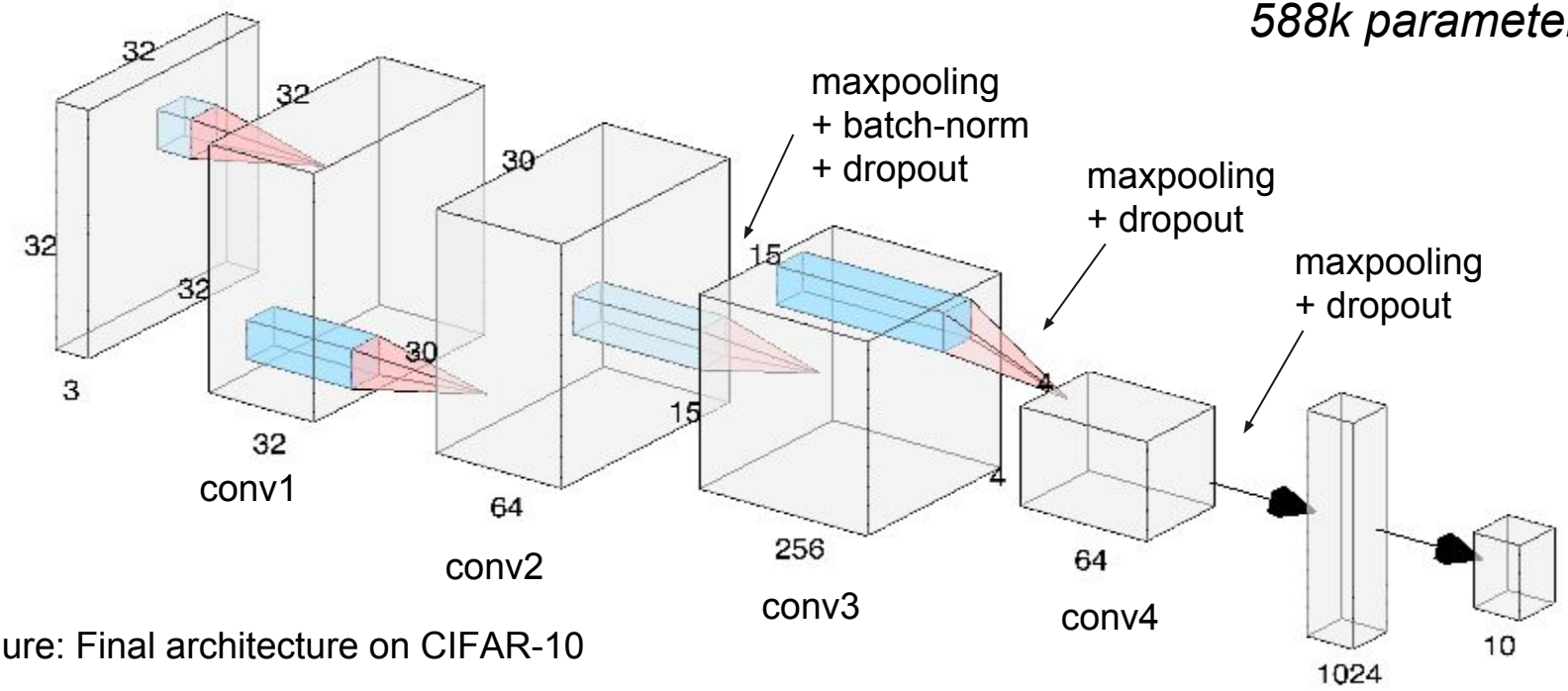

Figure: Final architecture on CIFAR-10

Sensitivity analysis

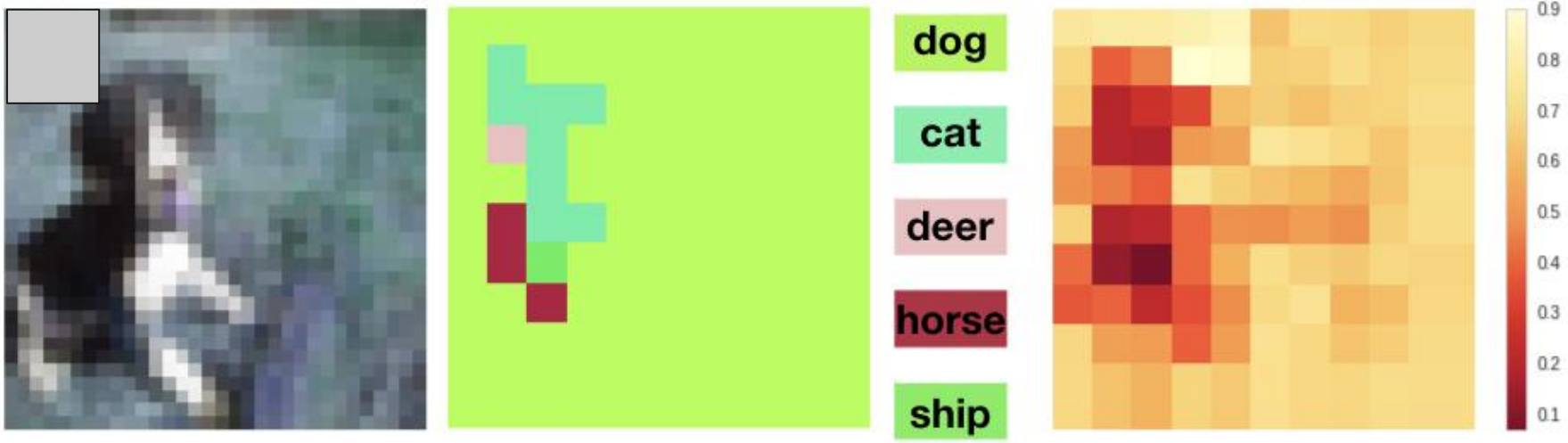

Figure: Sliding a gray patch on the input image