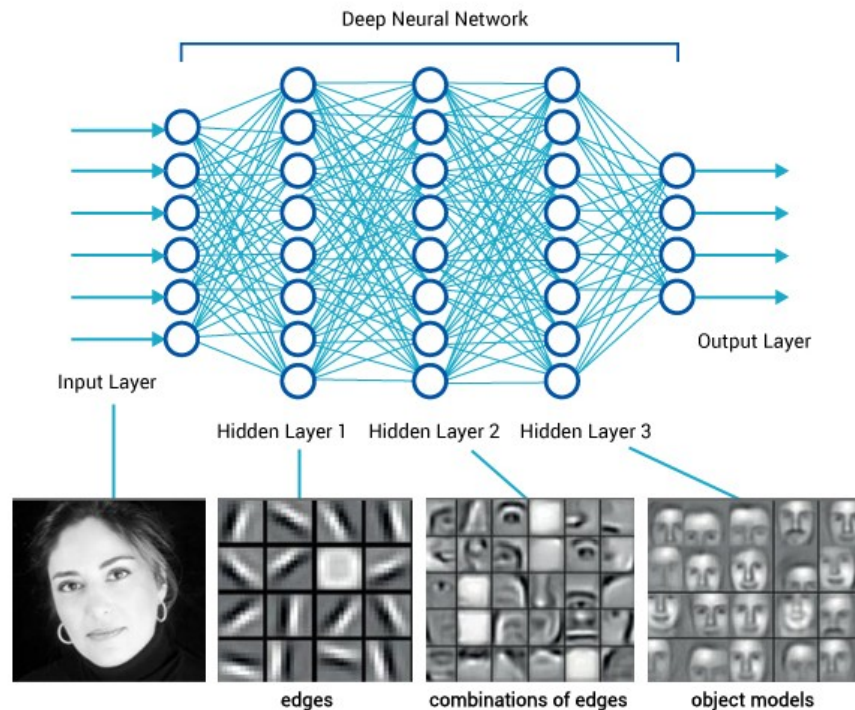


Machine learning

Lecture 6



Marcin Wolter
IFJ PAN

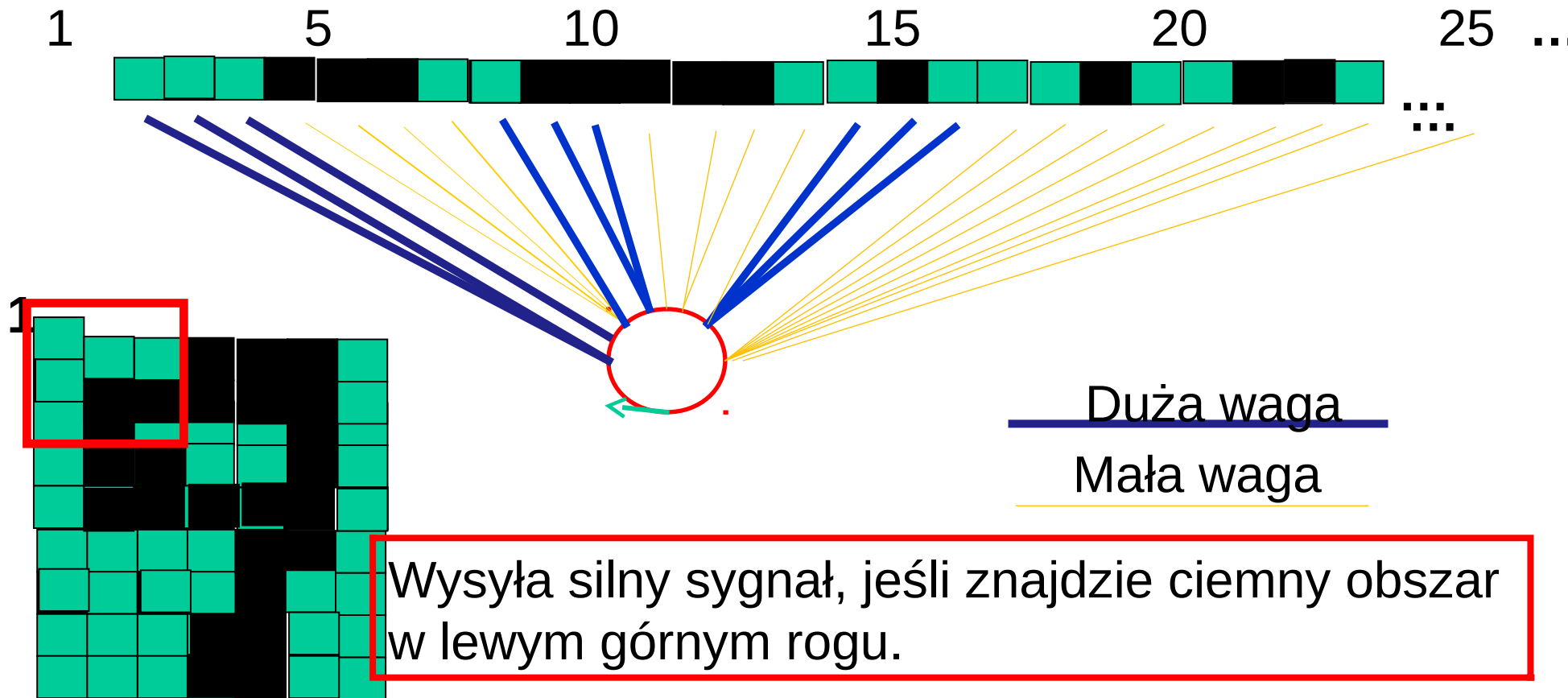
11 maja 2017

- Deep learning
- Convolution network
- Zastosowanie do poszukiwań bozonu Higgsa

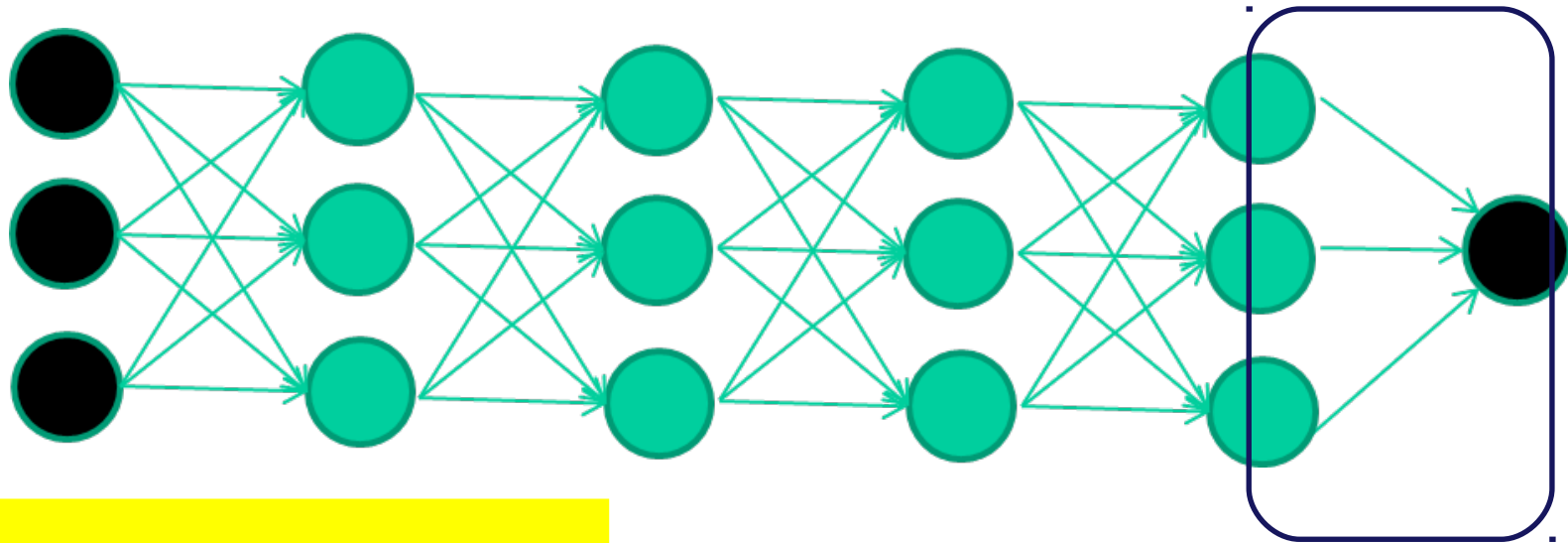


Deep learning

Poszczególne warstwy ukryte uczą się rozpoznawania „cech” - od najprostszycch do coraz bardziej złożonych.



Warstwy trenowane są sekwencyjnie



Trenuj tę warstwę

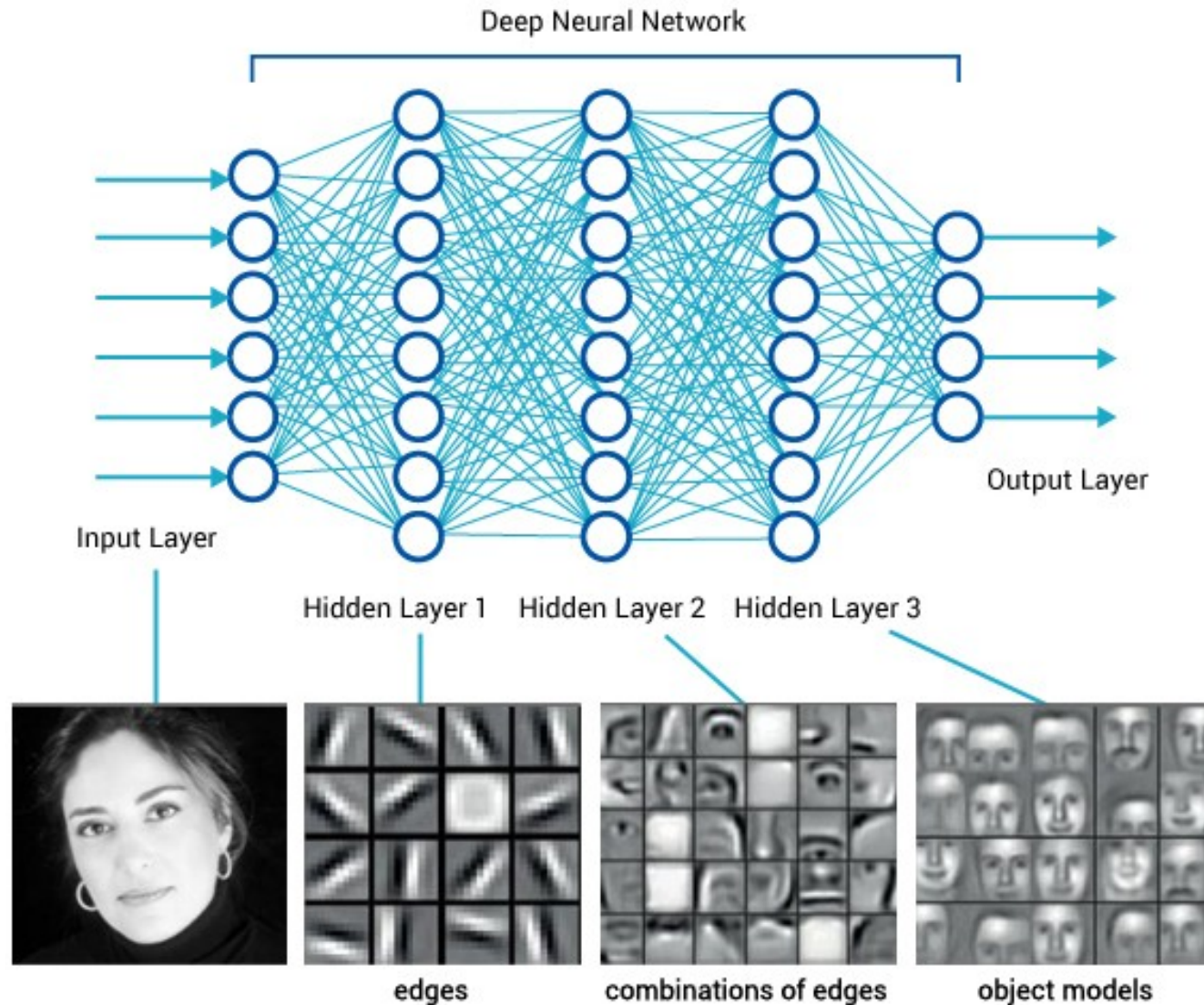
Potem ta

Potem ta

Potem ta

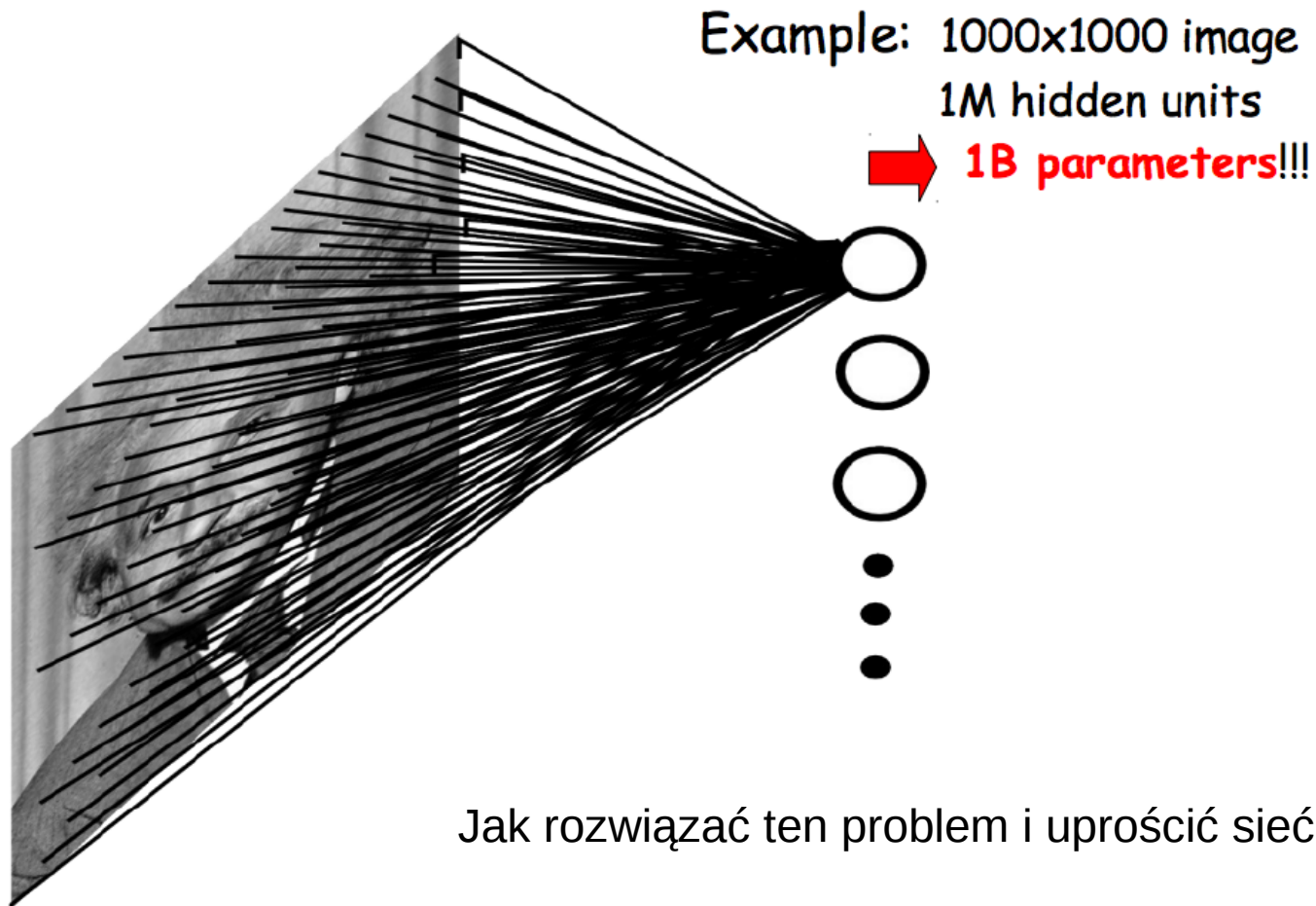
Na koniec ta

Deep Neural Network działa mniej więcej tak...

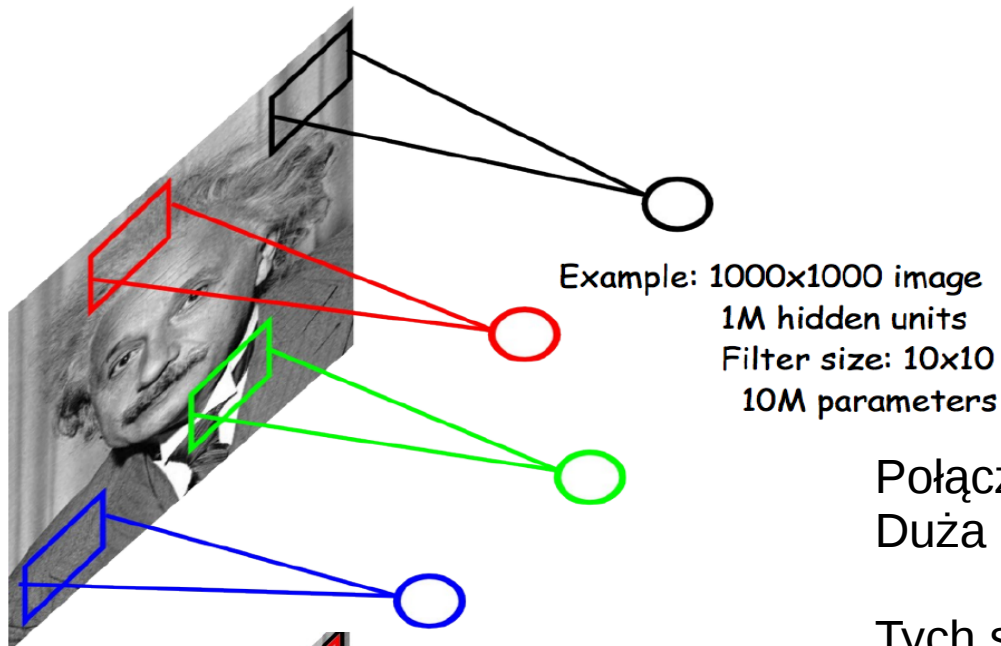


Convolutional NN

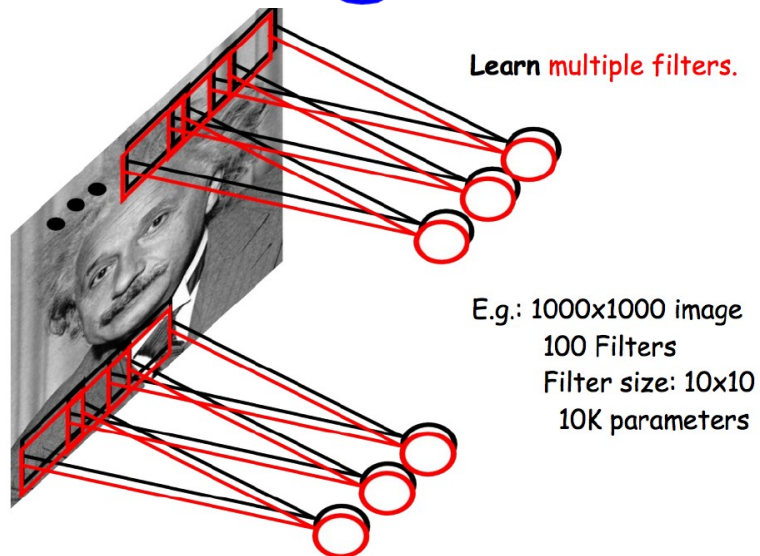
Rozpoznawanie obrazu



Convolutional NN



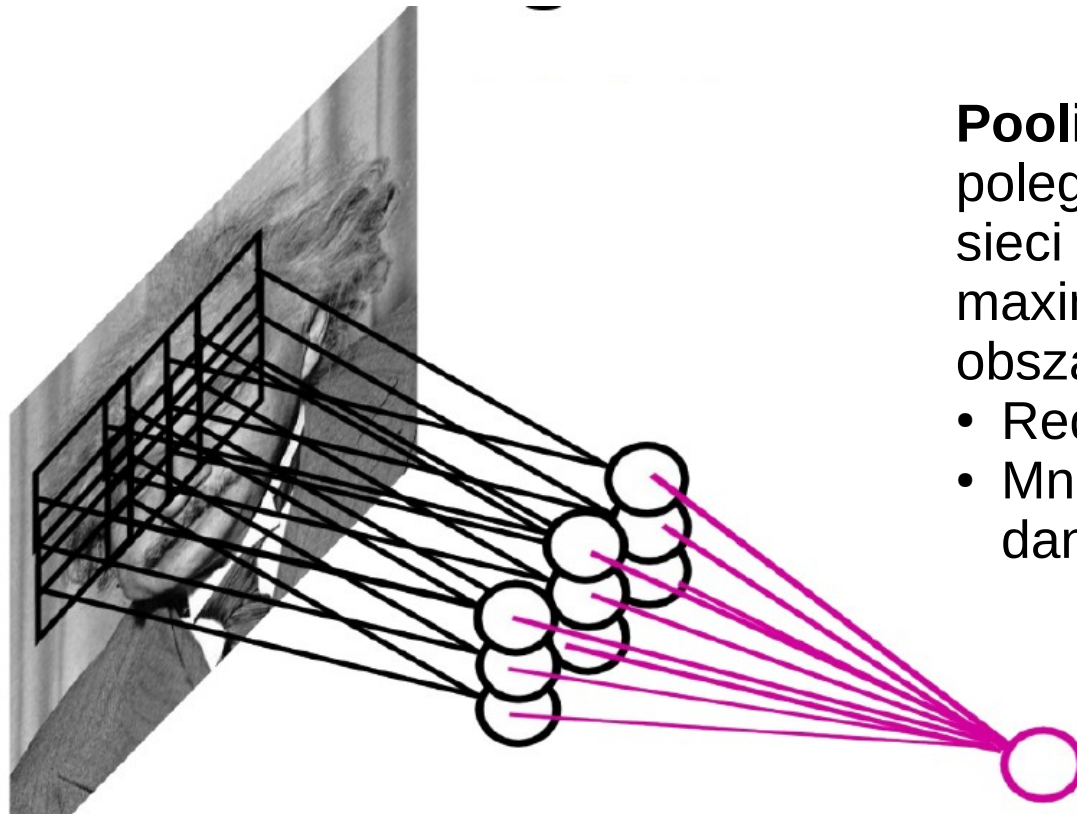
Połączyć tylko lokalne obszary, np. 10x10 pikseli
Duża redukcja liczby parametrów!



Tych samych cech można poszukiwać w różnych miejscach => wytrenujemy wiele różnych filtrów cech i przesuwamy je po obrazie.

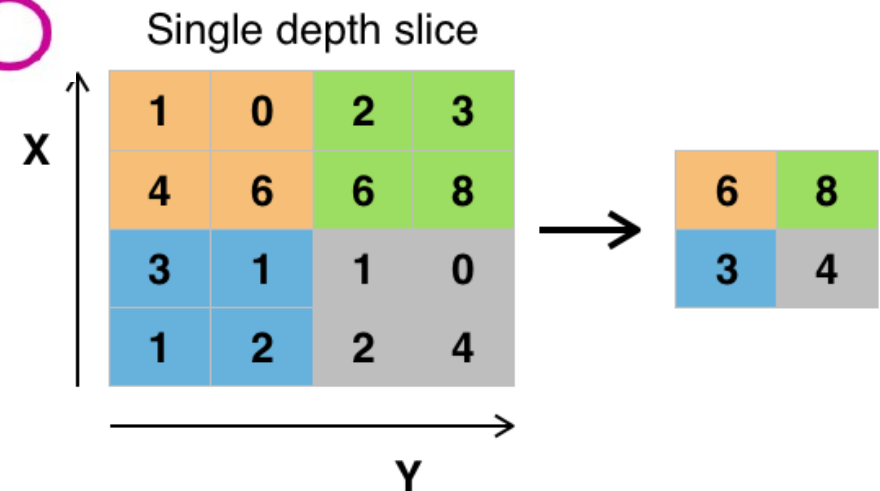
LeCun et al. "Gradient-based learning applied to document recognition" IEEE 1998

Pooling



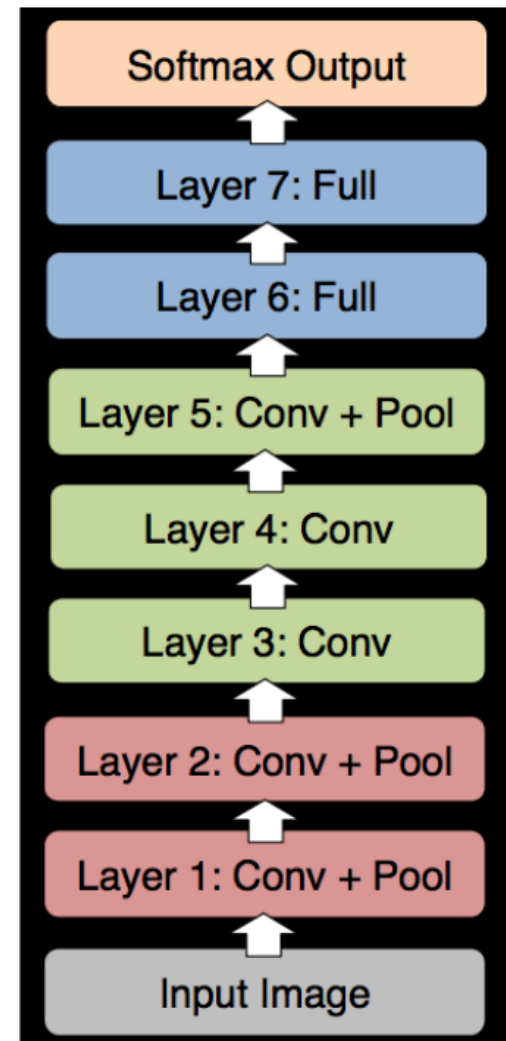
Pooling – (przeważnie **max pooling**) polega na zastępowaniu odpowiedzi sieci dla każdej pozycji np. przez maximum (albo średnią) dla danego obszaru:

- Redukcja danych
- Mniejsza wrażliwość na położenie danej cechy.



Architecture of Alex Krizhevsky et al.

- 8 layers total.
- Trained on Imagenet Dataset (1000 categories, 1.2M training images, 150k test images)
- 18.2% top-5 error
 - Winner of the ILSVRC-2012 challenge.



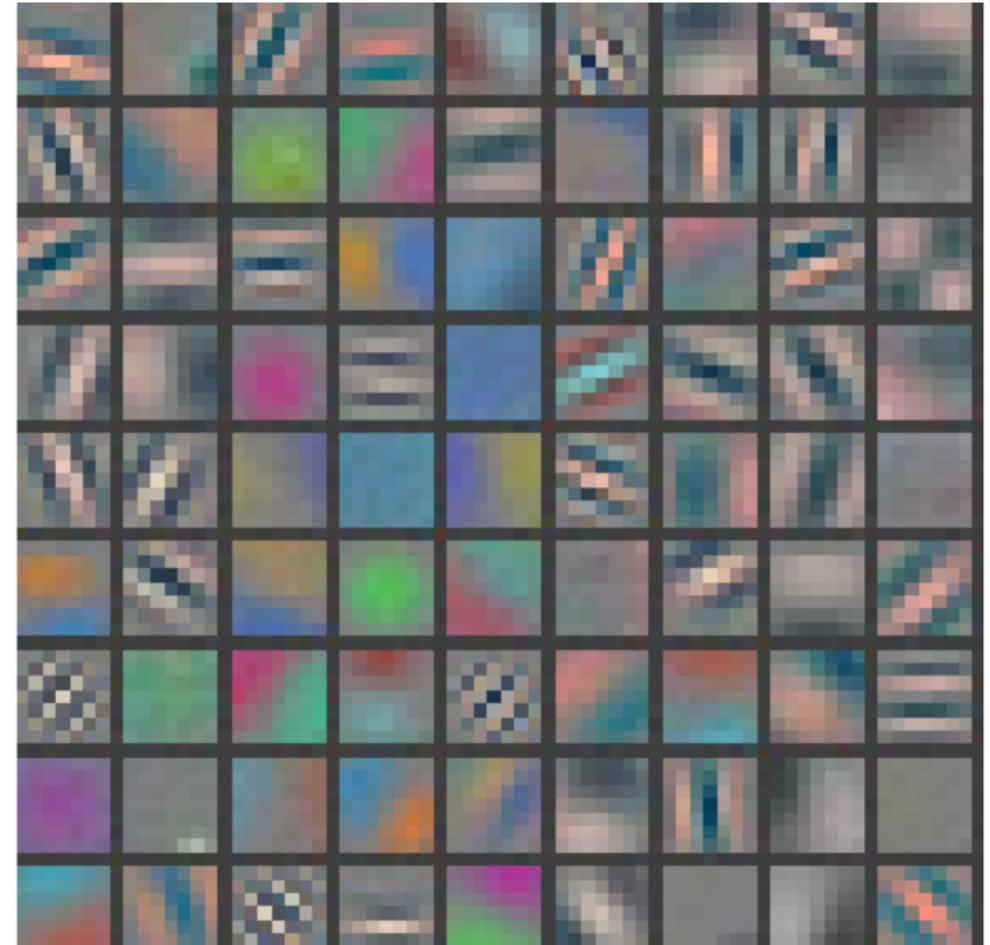
Slide: R. Fergus

First layer filters

Showing 81 filters of 11x11x3.

Capture low-level features like oriented edges, blobs.

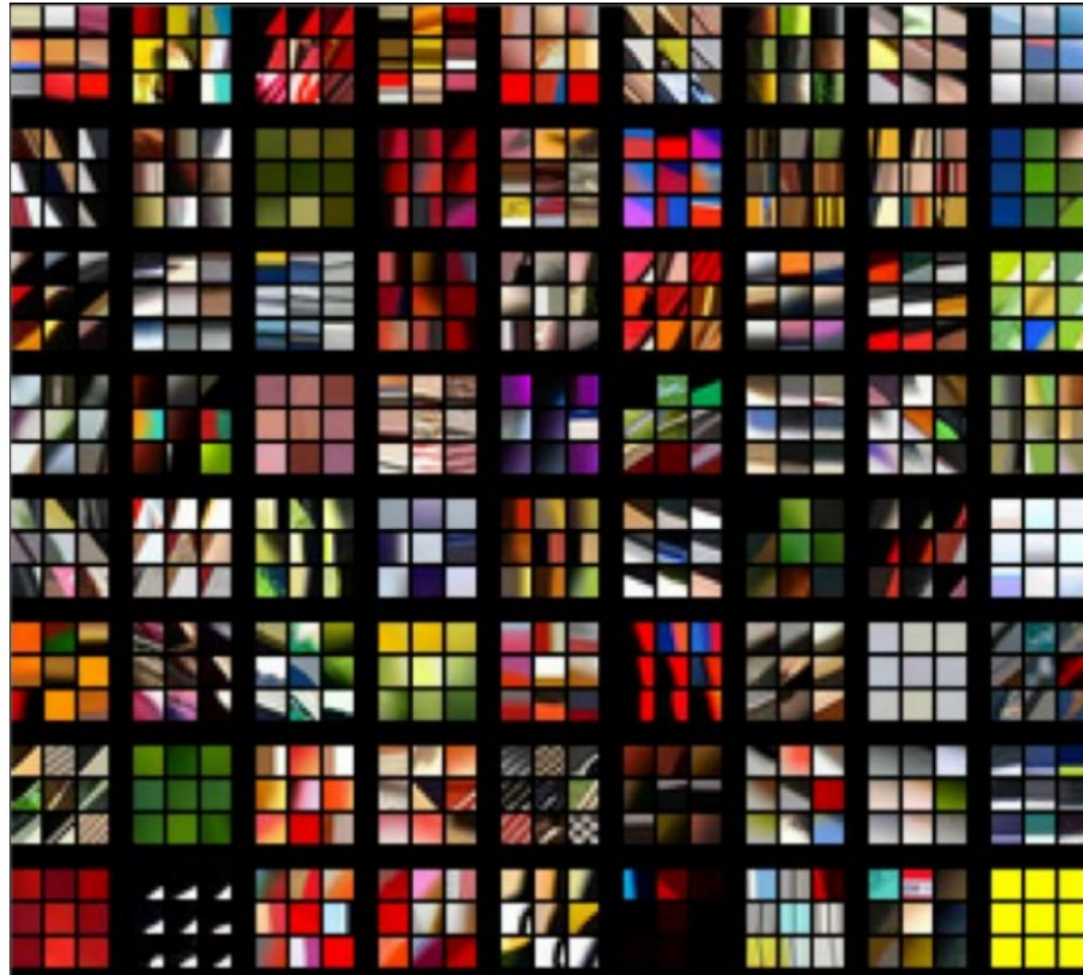
Note these oriented edges are analogous to what **SIFT** uses to compute the gradients.



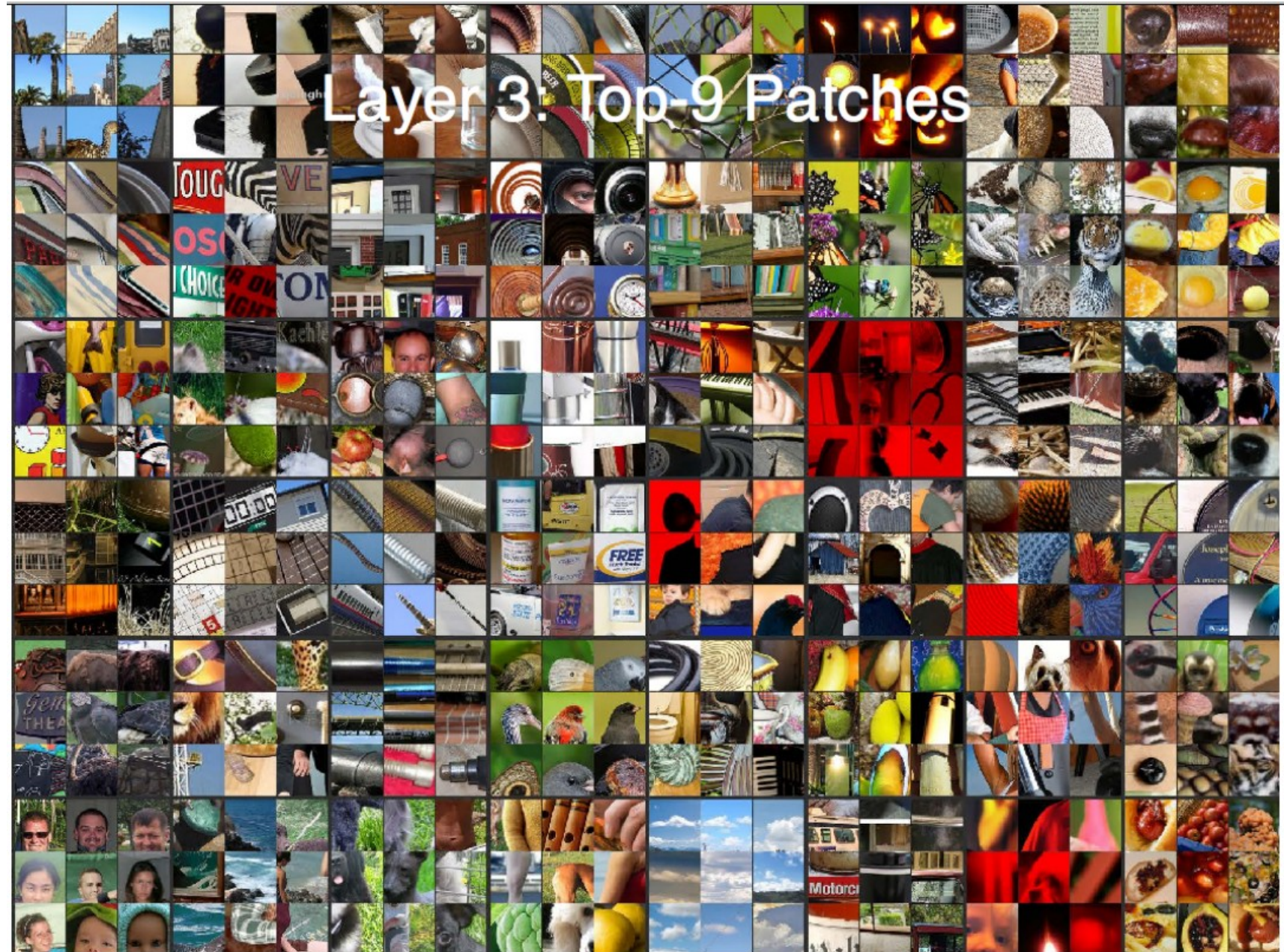
SIFT - scale-invariant feature transform, algorytm opublikowany w 1999 roku przez Davida Lowe'a.

Top 9 patches that activate each filter in layer 1

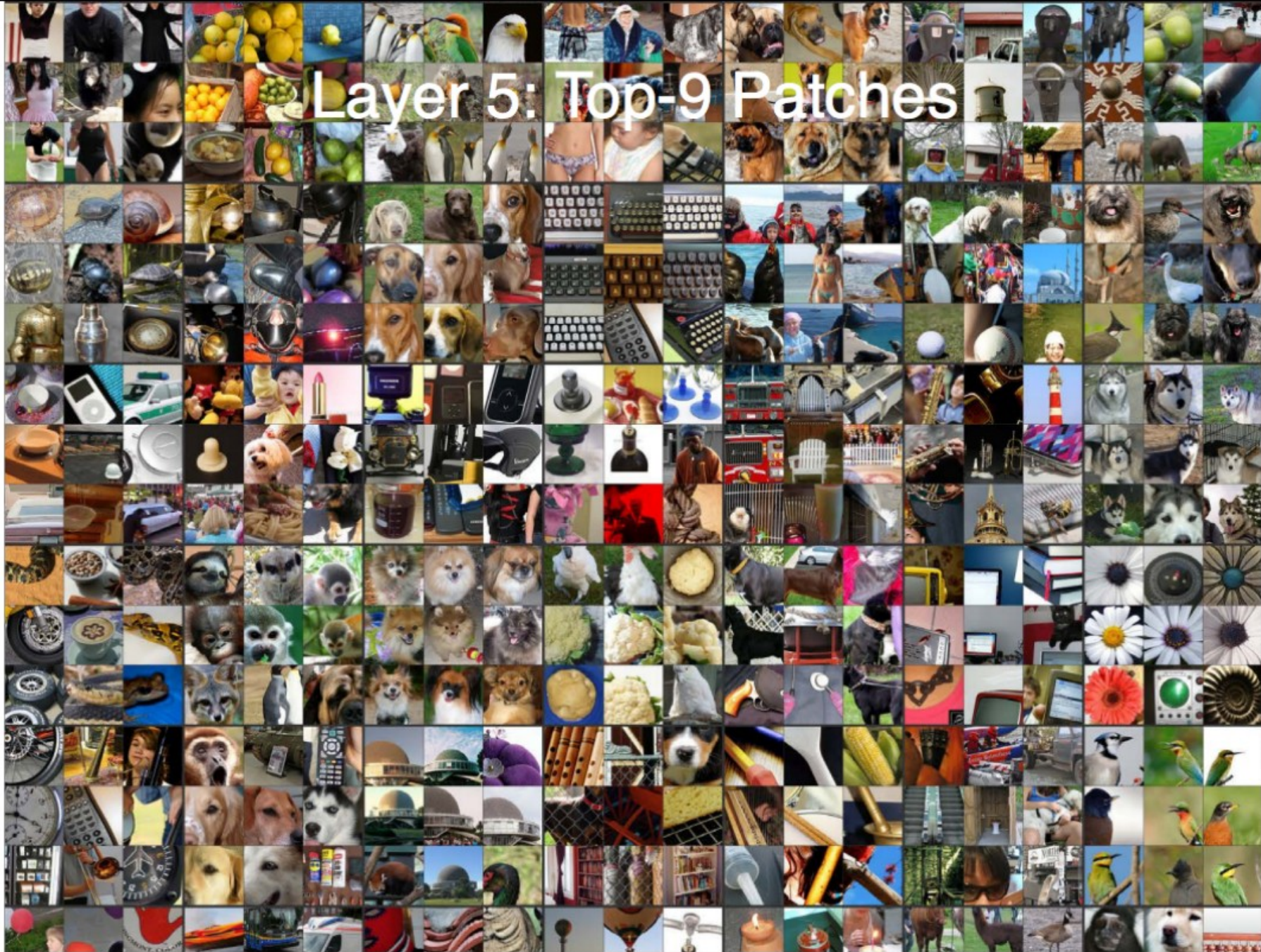
Each 3x3 block shows the top 9 patches for one filter.













Oprogramowanie

- **Biblioteki ConvNet:**

 - Cuda-convnet (Alex Krizhevsky, Google)

 - Caffe (Y. Jia, Berkeley, now Google)

 - Torch7 (Idiap, NYU, NEC)

- Scikit Learn (scikit-learn.org) również umożliwia użycie ConvNets

Kilka własności głębokich sieci

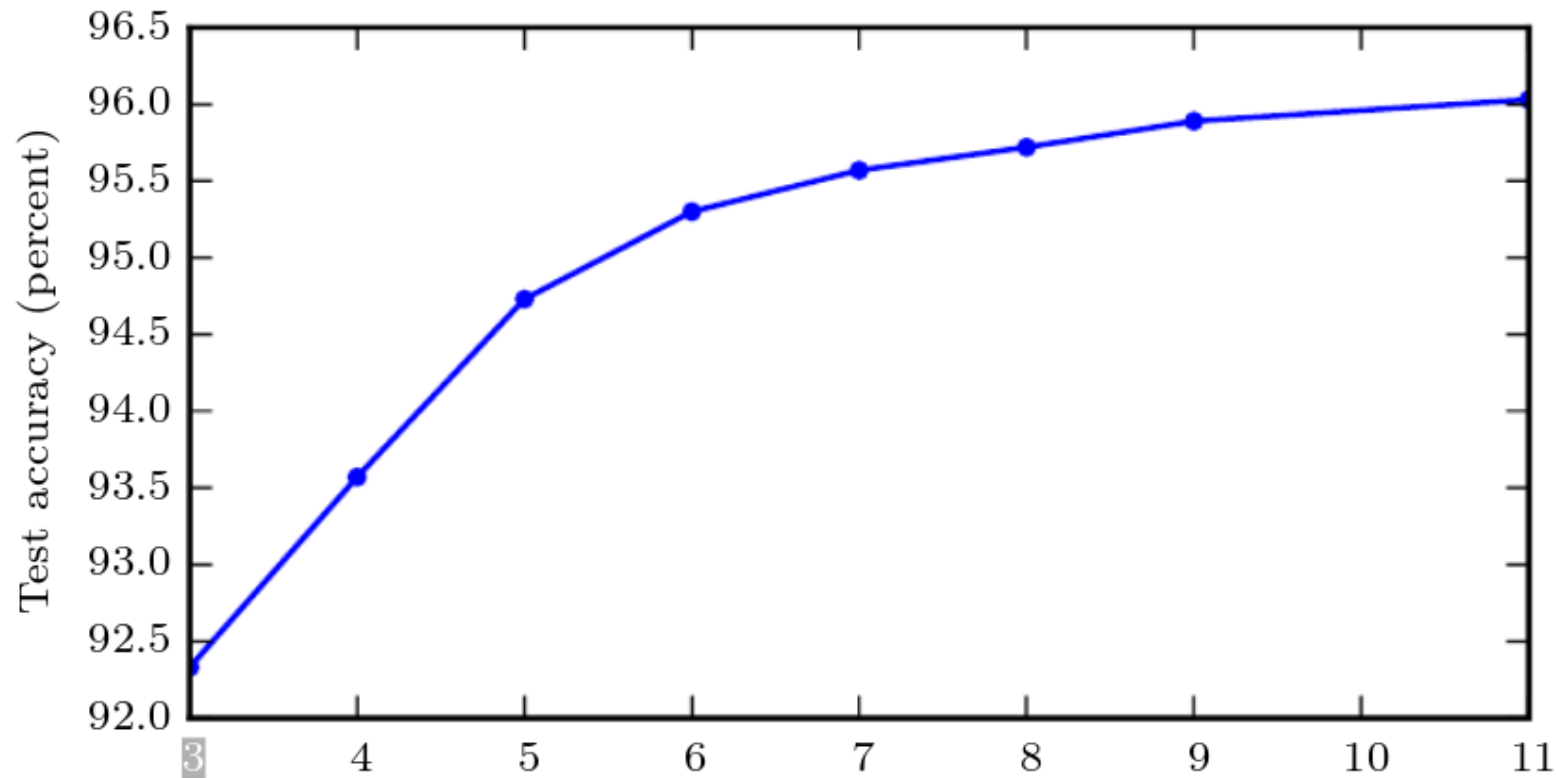


Figure 6.6: Empirical results showing that deeper networks generalize better when used to transcribe multi-digit numbers from photographs of addresses. Data from [Goodfellow et al. \(2014d\)](#). The test set accuracy consistently increases with increasing depth. See figure 6.7 for a control experiment demonstrating that other increases to the model size do not yield the same effect.

<http://www.deeplearningbook.org>

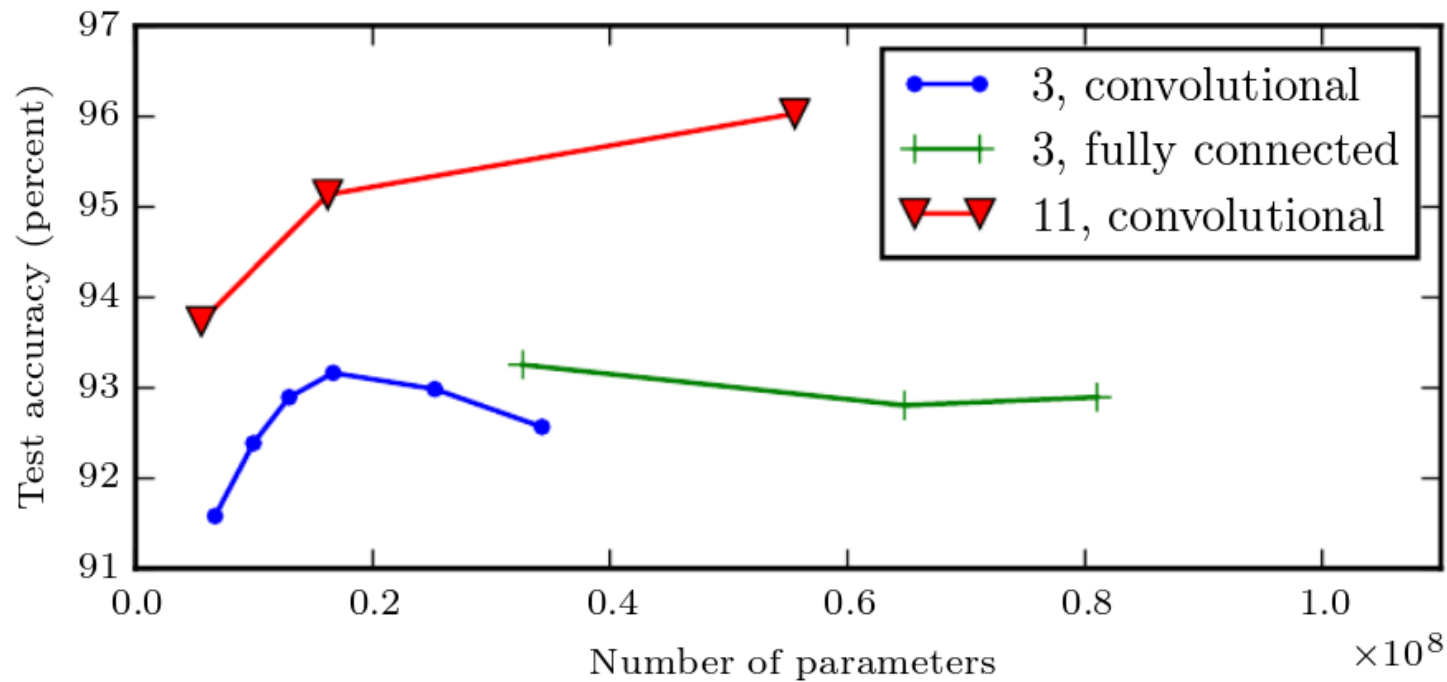
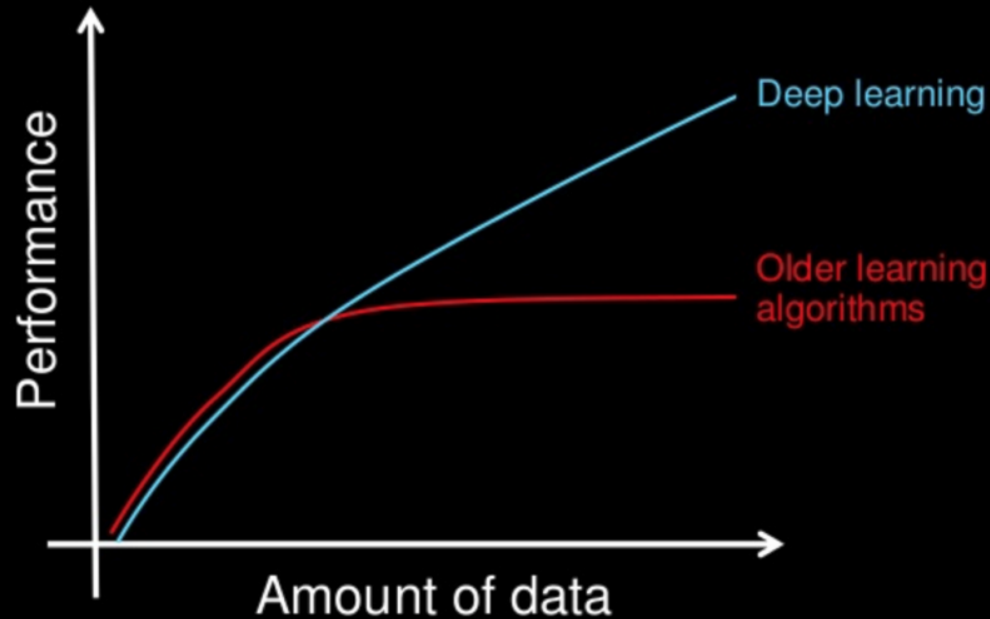


Figure 6.7: Deeper models tend to perform better. This is not merely because the model is larger. This experiment from [Goodfellow *et al.* \(2014d\)](#) shows that increasing the number of parameters in layers of convolutional networks without increasing their depth is not nearly as effective at increasing test set performance. The legend indicates the depth of network used to make each curve and whether the curve represents variation in the size of the convolutional or the fully connected layers. We observe that shallow models in this context overfit at around 20 million parameters while deep ones can benefit from having over 60 million. This suggests that using a deep model expresses a useful preference over the space of functions the model can learn. Specifically, it expresses a belief that the function should consist of many simpler functions composed together. This could result either in learning a representation that is composed in turn of simpler representations (e.g., corners defined in terms of edges) or in learning a program with sequentially dependent steps (e.g., first locate a set of objects, then segment them from each other, then recognize them).

Deep learning

Why deep learning



How do data science techniques scale with amount of data?

- Pięknie skaluje się przy dużej liczbie danych.

ARTICLE

Received 19 Feb 2014 | Accepted 4 Jun 2014 | Published 2 Jul 2014

DOI: 10.1038/ncomms5308

Searching for exotic particles in high-energy physics with deep learning

P. Baldi¹, P. Sadowski¹ & D. Whiteson²

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on 'shallow' machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.

Poszukiwanie bozonu Higgsa (symulowane dane CMS)

- Zmienne niskiego poziomu – 22 zmienne (tu pokazujemy kilka)

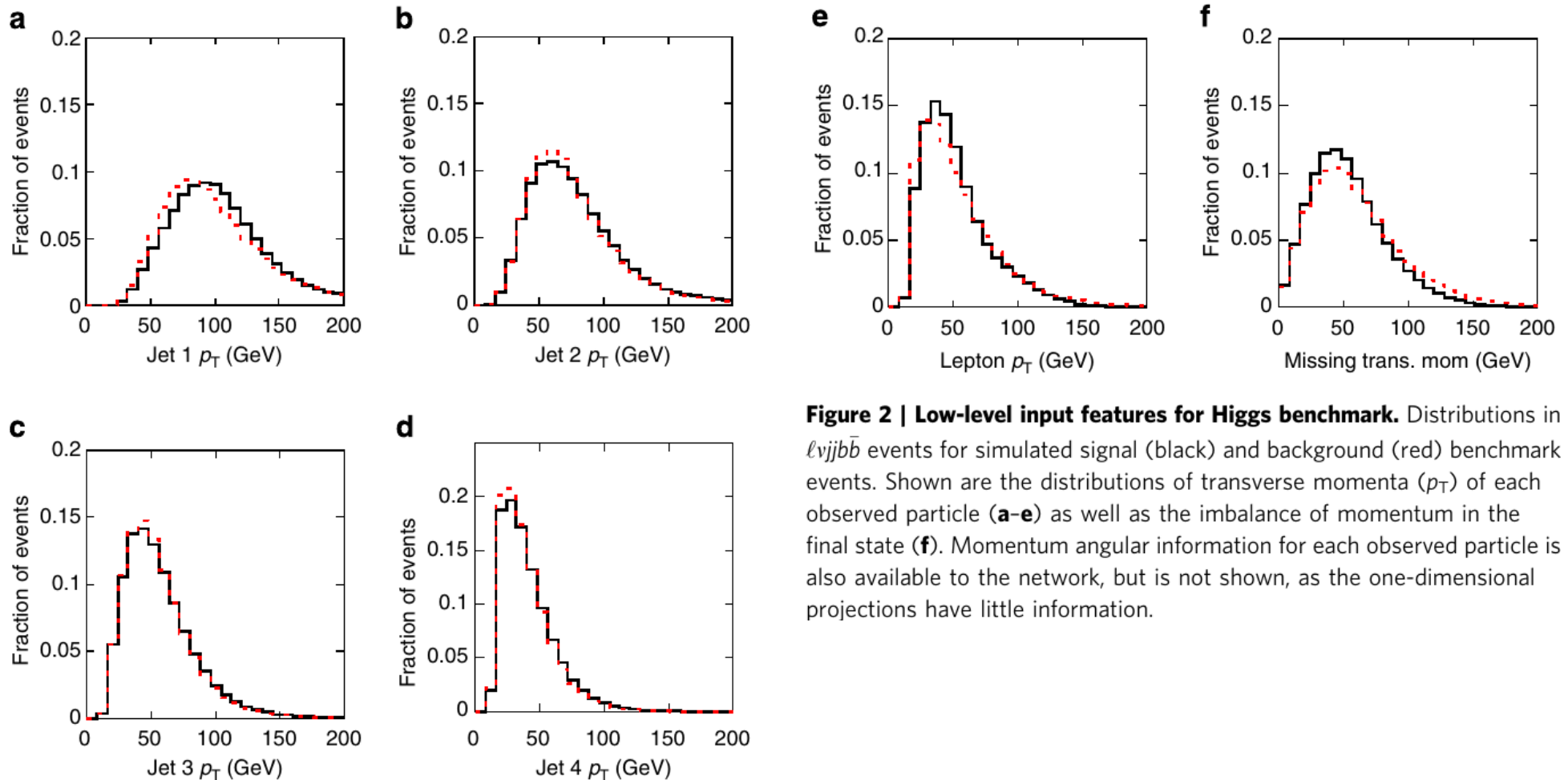


Figure 2 | Low-level input features for Higgs benchmark. Distributions in $\ell\nu jj\bar{b}\bar{b}$ events for simulated signal (black) and background (red) benchmark events. Shown are the distributions of transverse momenta (p_T) of each observed particle (**a-e**) as well as the imbalance of momentum in the final state (**f**). Momentum angular information for each observed particle is also available to the network, but is not shown, as the one-dimensional projections have little information.

Poszukiwanie Higgsa

- Fizycy pracują na dobrze dyskryminujących zmiennych wysokiego poziomu – są one zbudowane na podstawie zmiennych niskiego poziomu i nie zawierają żadnych dodatkowych informacji (7 zmiennych).

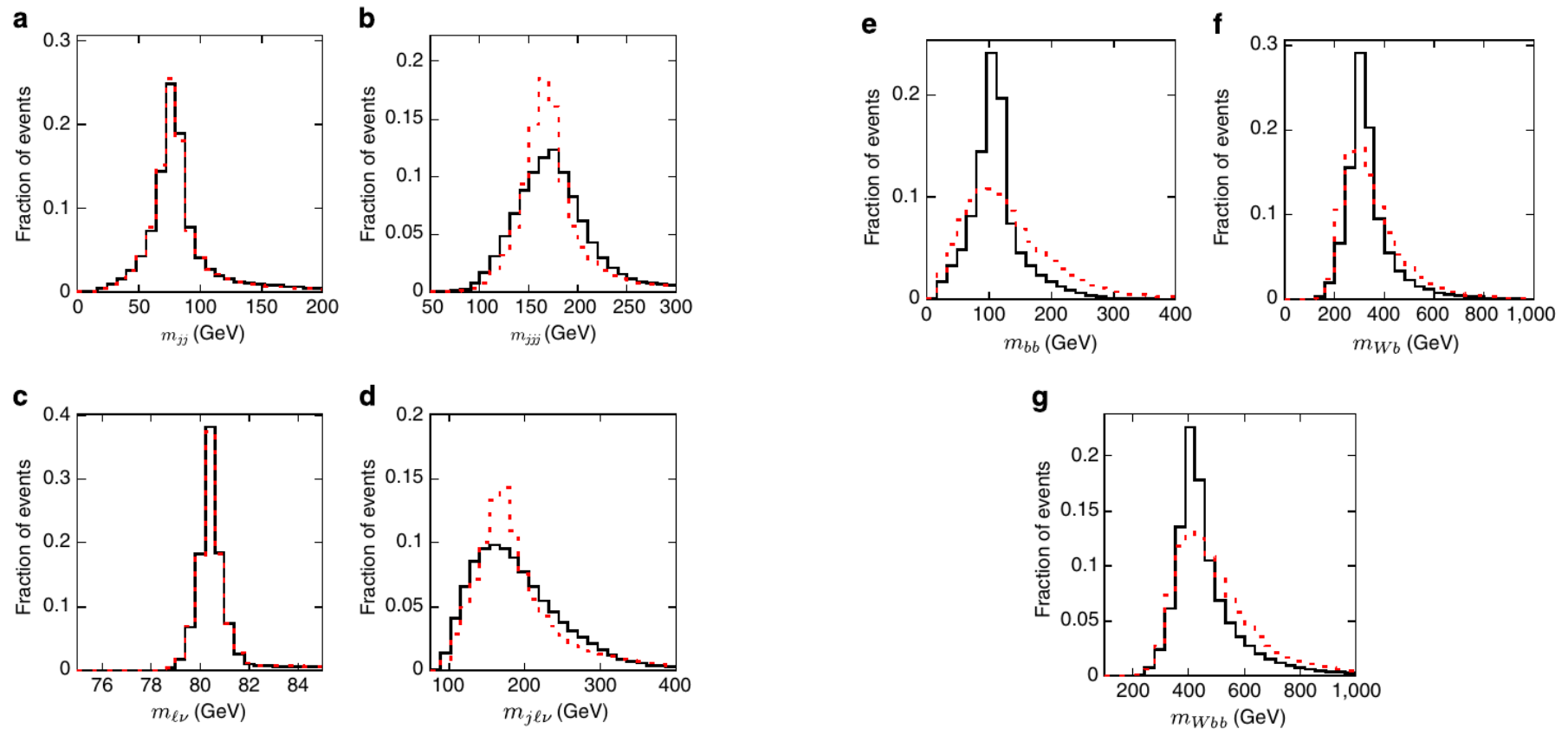


Figure 3 | High-level input features for Higgs benchmark. Distributions in simulation of invariant mass calculations in $\ell\nu jjb\bar{b}$ events for simulated signal (black) and background (red) events.

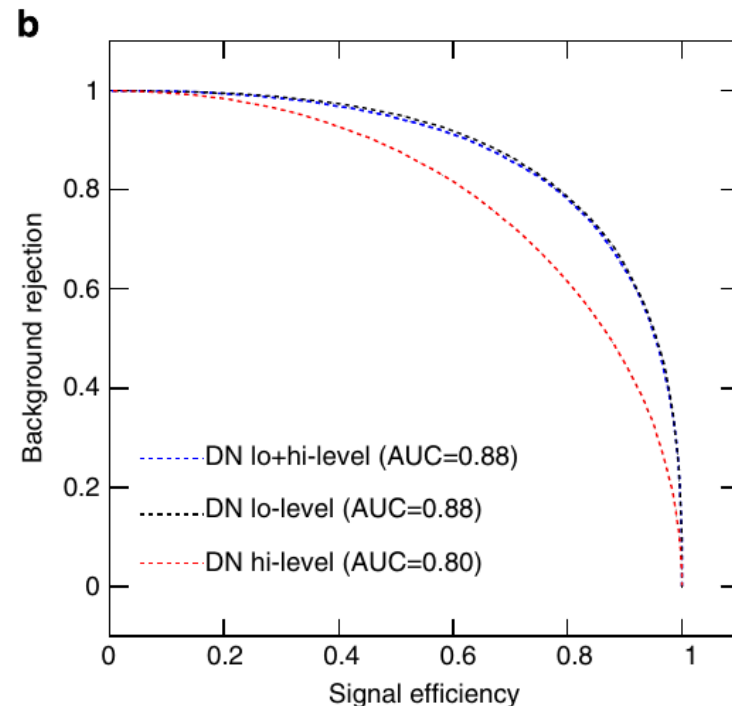
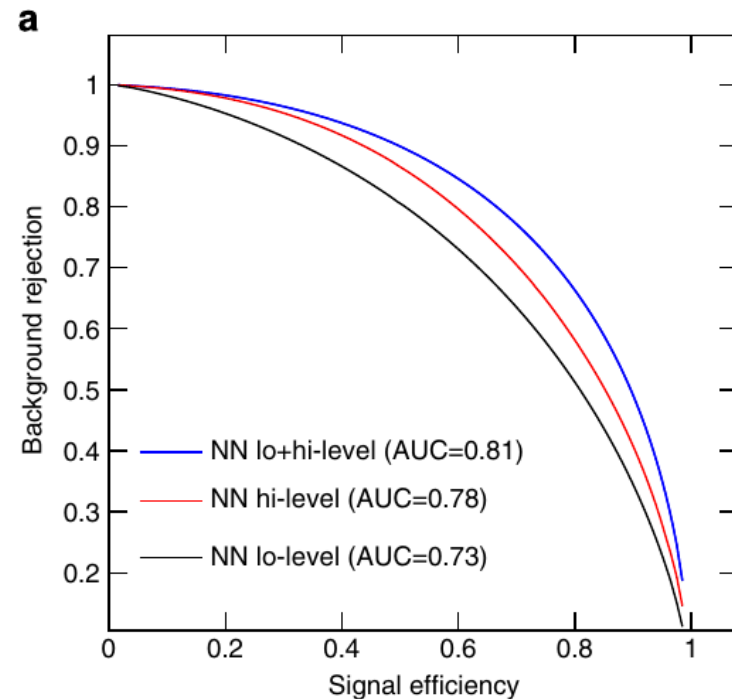
Poszukiwanie Higgosa

- Dane: 2 600 000 przypadków do nauki, 100 000 do walidacji
- Deep NN: 5 warstw, po 300 elementów w każdej

Table 1 | Performance for Higgs benchmark.

Technique	Low-level	High-level	Complete
<i>AUC</i>			
BDT	0.73 (0.01)	0.78 (0.01)	0.81 (0.01)
NN	0.733 (0.007)	0.777 (0.001)	0.816 (0.004)
DN	0.880 (0.001)	0.800 (<0.001)	0.885 (0.002)
<i>Discovery significance</i>			
NN	2.5 σ	3.1 σ	3.7 σ
DN	4.9 σ	3.6 σ	5.0 σ

Comparison of the performance of several learning techniques: boosted decision trees (BDT), shallow neural networks (NN), and deep neural networks (DN) for three sets of input features: low-level features, high-level features and the complete set of features. Each neural network was trained five times with different random initializations. The table displays the mean area under the curve (AUC) of the signal-rejection curve in Fig. 7, with s.d. in parentheses as well as the expected significance of a discovery (in units of Gaussian σ) for 100 signal events and $1,000 \pm 50$ background events.





Podsumowanie

- Deep NN znalazła cechy umożliwiające znalezienie bozonu Higgsa lepiej, niż fizycy.
- Może pozwoli to w przyszłości na automatyzację analiz fizycznych?
- Czyżby groziło nam bezrobocie???



Zadanie zaliczeniowe



ATLAS Z → tau tau selection

- Dane:

- mc12/Ztautau.root - sygnał
- Powheg_ttbar.root - tło
- Wenu.root - tło
- Wmunu.root - tło
- Wtaunu.root - tło
- Zee.root - tło
- Zmumu.root - tło

- Zmienne:

preselekcja:

```
if(!(      evtSel_is_dilepVeto > 0 && evtSel_is_tau > 0 &&
fabs(evtSel_tau_eta) < 2.47 && evtSel_is_conf_lep_veto == 1 &&
evtSel_tau_numTrack == 1 && evtSel_lep_pt > 26 &&
fabs(evtSel_lep_eta) < 2.4 && evtSel_transverseMass < 70))
continue;
```

```
if (!( evtSel_is_oppositeSign>0 && evtSel_is_mu>0 &&
evtSel_is_isoLep>0 )) continue;
```



ATLAS $Z \rightarrow \tau\tau$ selection

- Zmienne użyte do treningu:
 - *evtsel_tau_et*
 - *evtsel_dPhiSum*
 - *evtsel_tau_pi0_n*
 - *evtsel_transverseMass*
 - *sum_cos_dphi*
- Spectator
 - *vis_mass*
- Program:
 - TMVAClassificationMW.C i TMVAClassificationMW.h
Wykonuje podstawowy trening.



ATLAS Z → tau tau selection

- Zainstalować pakiet root i TMVA
- Ściągnąć dane i przykładowy program:
 - <http://nz14-46.4.ifj.edu.pl/cwiczenieATLAS/>
- Uruchomić przykładowy program:

```
root -l
.L TMVAClassificationMW.C++
TMVAClassificationMW t
t.Loop()
```
- Zmodyfikować go:
 - **Spróbować zoptymalizować parametry wybranej metody**
 - **Spróbować usunąć jakieś zmienne a może dodać?**
 - **Spróbować użyć indywidualnych zmiennych wchodzących w skład np. *sum_cos_dphi***
 - **Użyć wszystkich rodzajów tła – użyć wag *WeightLumi***
- Zaaplikować wyuczony klasyfikator do danych (*data12/Muons.PhysCont.grp14.root*), można się wzorować na przykładzie *TMVAClassificationApplication* dostępnym na stronie TMVA oraz załączonym przykładzie *TMVAClassificationApplicationMW.C*.

ATLAS $Z \rightarrow \tau\tau$ selection

Wykonać tego typu rysunek np. dla masy widzialnej

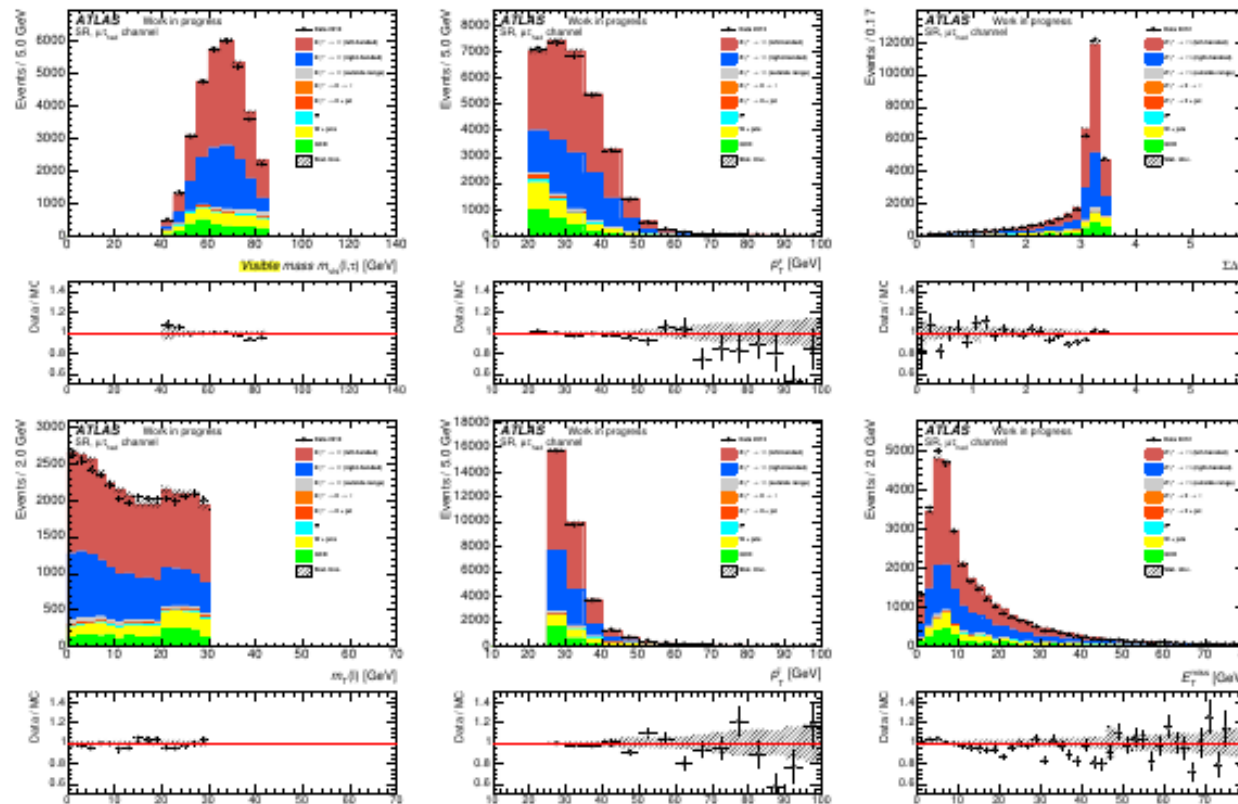


Figure 41: Distributions of variables observed in $Z \rightarrow \tau\tau$ (μ -had channel). From top-left: visible mass of τ -lepton system, τ transverse momentum, sum of polar angles between τ and missing- E_T and between lepton and missing- E_T , transverse mass of the lepton-missing- E_T system, lepton transverse momentum and missing- E_T .



ATLAS $Z \rightarrow \tau\tau$ selection

Event Selection and Background Estimate

Region / Cut	Signal Region	Same Sign	W Control Region	QCD Control Region
Single lepton trigger + offline lepton pT	evtsel_is_mu / evtsel_is_el			
Isolated Lepton	evtsel_is_isoLep			!evtsel_is_isoLep
Medium Tau ID	evtsel_is_tau			
Veto dileptons	evtsel_is_dilepVeto			
Muon Veto + medium Electron Veto	evtsel_is_conf_lep_veto_medium			
Single Prong tau	evtsel_tau_numTrack == 1			
Transverse Mass	evtsel_transverseMass < 30		evtsel_transverseMass > 70	
Sum Delta Phi	evtsel_dPhiSum < 3.5		evtsel_dPhiSum > 3.5	
Opposite Sign	evtsel_is_oppositeSign	!evtsel_is_oppositeSign	evtsel_is_oppositeSign / !evtsel_is_oppositeSign	evtsel_is_oppositeSign / !evtsel_is_oppositeSign

**Cięcia zastosowane w analizie polaryzacji tau pochodzących z rozpadu $Z \rightarrow \tau\tau$
Czy używając uczenia maszynowego udało nam się poprawić wynik?**