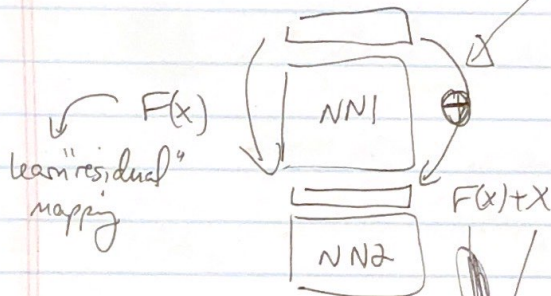


Improvements to basic CNN architecture

(2015) Highlight one: ResNet "skip connections"

enables much deeper NNs w/out vanishing/exploding gradients

e.g. 50, 100 layers!
150 (or even 1000 layers!)



Different blocks can learn different features without feeding into another
Can skip blocks

Can also put a weight in front of x

"gate"

but doesn't seem to be perf. gain, might even be worse!

might need zero padding or truncation/clipping to match dim. of $F(x)$

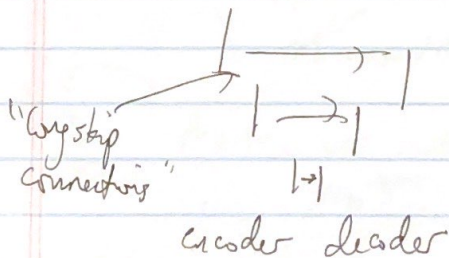
only works b/c some has identity mapping, close to optimal!!

Seems almost trivial but it dramatically improves performance!

• Also: DenseNet — concat instead of add

- Also: UNet used to map images to ~~themselves~~ other images

- denoising
- segmentation/obj detection



- Top tagging challenge — ResNet improved on simple CNN by 20-30% in bg rejection @ θ for $\approx 30\%$

Example 2: Galaxy Zoo Kaggle Competition

(From 2014 — predates ResNet!)

1308.3496

Based on Galaxy Zoo 2 dataset — 304k galaxies

images of
from SDSS

Classified by citizen scientists
60M classifications

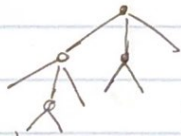
80,000 volunteers

Cont. part:
v. diff than
top tagging challenge
— no simulation
no ground truth

Decision tree

questionnaire w/ 11 Q's → 37 possible responses

(unique paths through the tree)



regression
not classification!

Same Galaxy classified by multiple people.

→ set prob of each response in questionnaire

e.g. smooth, disk, star/artifact

80% answered this 15% 10%

Goal is to
mimic citizen
scientists
→ AI scientist!

Metric was matching these probabilities w/ simple MSE

$$L = \sum_{i=1}^n \sum_{a=1}^m (p_{ia} - \hat{p}_{ia})^2$$

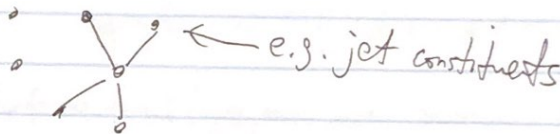
→ winner was a ~~simple~~ classic

Convnet trained for 67 hours!

More NN architectures beyond CNNs

Another popular framework, also permutation inv't:
graph neural networks "message passing NN"

View ^{each point} data as a graph — more flexible than image!



↓
 no discretization
 better of phase space
 images

graphs have nodes and edges

↑
 attributes (p_x, p_y, p_z) e.g.

Diff types of graphs: fully connected



Sets (no connections)

hierarchical trees

locally connected

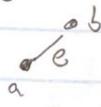
— each node connected

to nearest neighbors (by some dist metric)



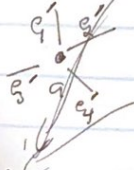
Message passing framework

∀ edge



$$\phi_{\text{edge}}(e, a, b) = e'$$

∀ node



$$\phi_{\text{node}}(\phi(e_1, e_2, \dots), a) = a'$$

$\phi_{\text{edge}}, \phi_{\text{node}} \rightarrow$ MLPs usually

Same \forall edge/node weight sharing like CNN filters!

Symmetric fn e.g. sum, mean, max
 ↓
 perm inv't