The Neural Tangent Link Between CNN Denoisers and Non-Local Filters

Julián Tachella, Junqi Tang and Mike Davies

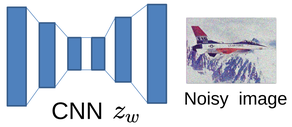
THE UNIVERSITY of EDINBURGH

CVPR VIRTUAL JUNE 19-25

Understanding CNN denoisers

- How do CNNs perform denoising?
- Implicit learning bias?
- Deep image prior phenomenon [1]

$$\arg\min_w \|z_w - y\|^2$$



1. Only access a noisy image y
2. Trainable weights w >> training pixels in y
3. Early stopping provides very good denoising?!

Neural tangent kernel (NTK) perspective

Assuming:

1. Overparameterization: channels of hidden layers $c \to \infty$

2. Standard iid **initialization** of weights *(e.g., He, LeCun)*

3. Gradient descent with learning rate $\eta \propto c^{-1}$

- Weights move little from initialization $\|w^t - w^0\|_2 \approx \mathcal{O}(c^{-0.5})$
- The network's output dynamics are given in closed form by

$$z^{t+1} = z^t + \eta\Theta(y - z^t)$$

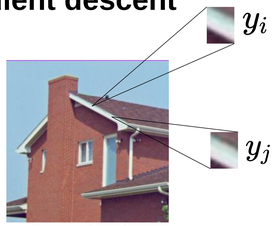
where the NTK [2] $\eta\Theta = \eta \frac{\delta z}{\delta w}(\frac{\delta z}{\delta w})^\top\Big|_{w=w^0}$

- Stays **constant** throughout training
- Available in **closed form**
- Only **depends on network's input** and architecture

1. Input noisy image + gradient descent

Patch similarity function

$$[\eta\Theta]_{i,j} = k(y_i, y_j)$$

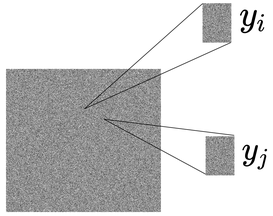


Input noisy image

- NTK = **non-local filter**
- CNN optimization = **iterative filtering** (twicing)
- Early stopping controls bias-variance tradeoff
- CNNs can be replaced by **closed form NTK** + Nyström

2. Input iid noise + Adam

- Deep image prior setting
- NTK = simple **low-pass filter**



Input iid noise

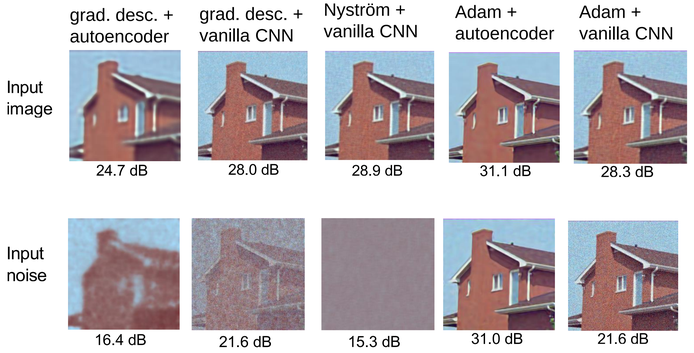
Fixed NTK **cannot explain** good performance!

- Weights **move significantly** from initialization $\|w^t - w^0\|_2 \approx \mathcal{O}(1)$
- Tangent kernel adapts throughout training:

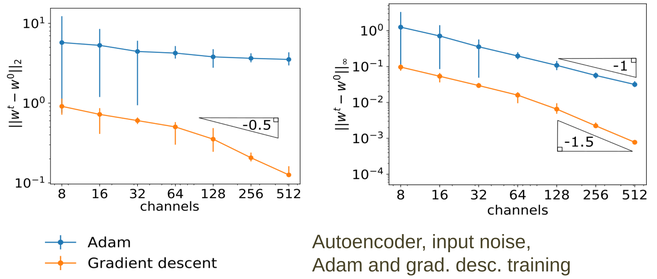
$$z^{t+1} = z^t + \eta\Theta^t(y - z^t)$$

Experiments:

Denoising performance



Change of weights throughout training



Autoencoder, input noise,
Adam and grad. desc. training

Summary

- CNNs posses a very strong bias towards clean images
- NTK gives the associated patch similarity function
- Effective degrees of freedom ≪ parameters in the network
- Optimizer plays key role in implicit bias

[1] Ulyanov et al. "*Deep image prior*", CVPR 2018

[2] Jacot et al. "*Neural tangent kernel: Convergence and generalization in neural networks*", NIPS 2018

project website here