Object Detection using the Scattering Transform

Marius Hobbhahn

July 14, 2019

Motivation

Problems

1. Loads of data necessary for training
2. Capability to generalize unclear for different circumstances (i.e. equivariances and invariances w.r.t. some transformations)

Possible solutions

1. Filters that generalize quickly
2. Filters that are globally equivariant and locally invariant w.r.t. some transformations, i.e. translation, rotation, scale

Scattering Transform

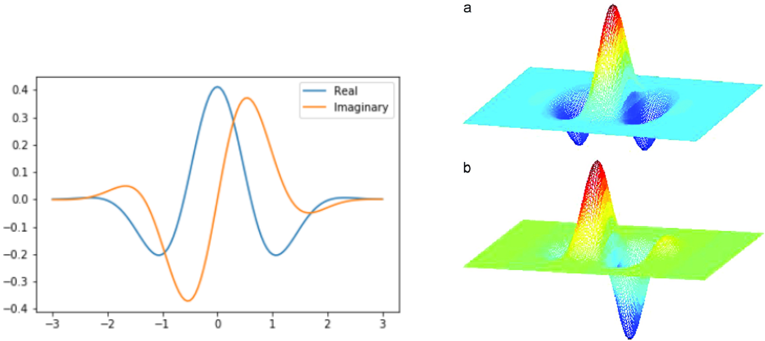
Basic Idea

Static image filter that has certain theoretical guarantees with respect to global equivariances and local invariances (i.e. translation, scale, rotation).

$$\psi(u) = C_1(e^{iu.\xi} - C_2)e^{\frac{-|u|^2}{2\sigma^2}} \tag{1}$$

- ξ: central frequency ($3\pi/4$)
- σ: width of the Gaussian part (0.85)
- C_1, C_2: Constants, C_2 is chosen s.t. $\int \psi(u)du = 0$ and $C_1 = 1$

Visualization of the Scattering Transform



(A) (B)

Figure 1: Complex Morlet wavelet in 1D and 2D

Visualization of the filter bank

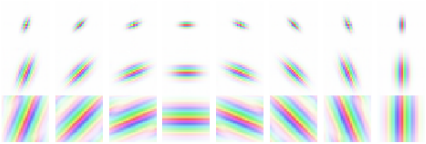


Figure 2: Visualization of the filter bank. $j = 3$ is denotes the down scale factor and $\theta = 8$ the number of angles. Color saturation and color hue respectively denote complex magnitude and complex phase.

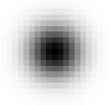
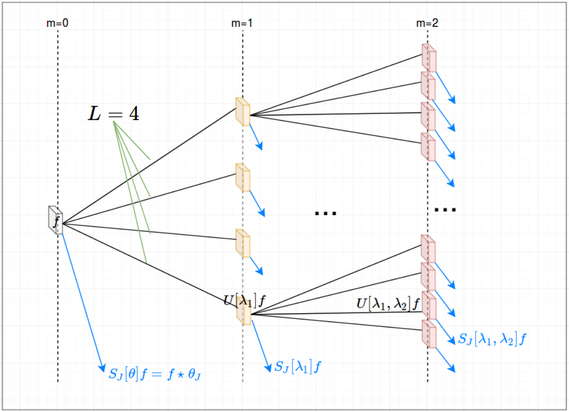


Figure 3: Visualization of the low pass (Gaussian) Filter

Scattering Networks

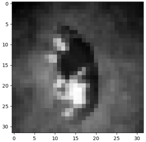
Basic Idea

Apply the scattering transform multiple times to get higher order scattering coefficients.

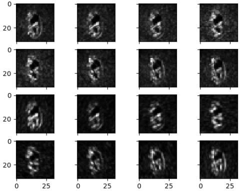


Example

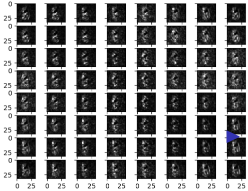
(A)



(B)



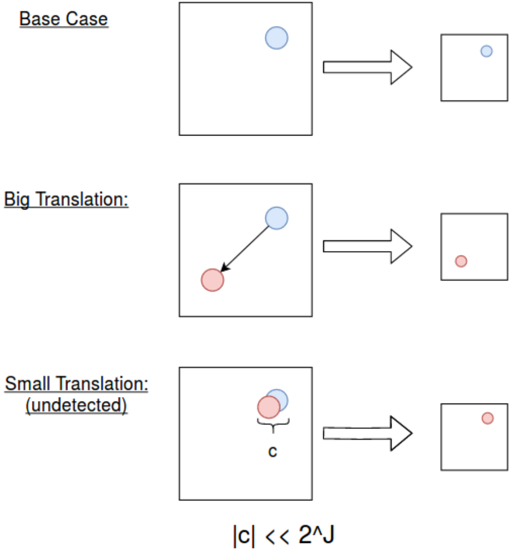
(C)



(D)

- Example for $L = 8$, $J = 2$, $N, M = 128$
- a) original image
- b) Gaussian low-pass filter
- c) first order scattering coefficients (size 32x32)
- d) second order scattering coefficients (size 32x32)

Properties of the Scattering Transform



- Invariance: $f(Tx) = f(x)$
- Equivariance: $f(Tx) = Tf(x)$
- Local invariance but global equivariance

Hybrid scattering networks for classification

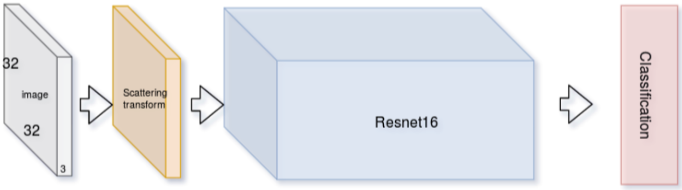
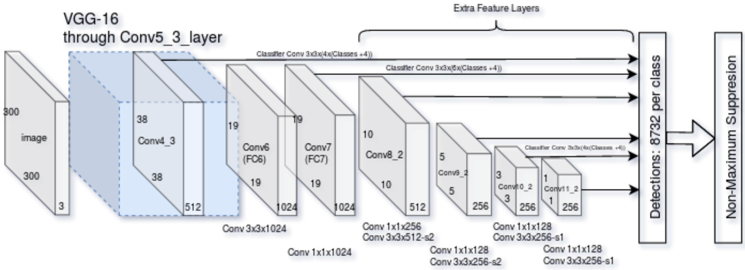


Figure 5: Architecture [OBZ17]

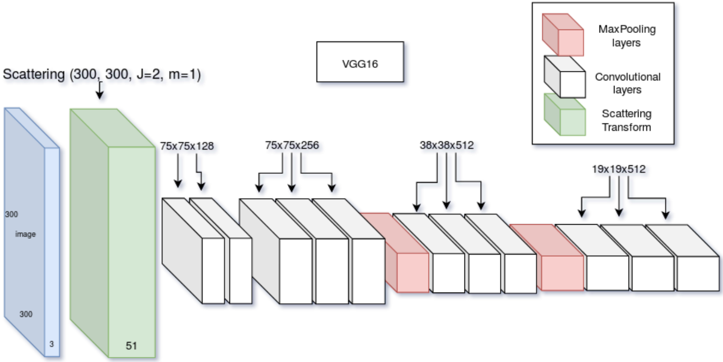
Method	100	500	1000
WRN 16-8	34.7 $\pm$ 0.8	46.5 $\pm$1.4	60.0 $\pm$1.8
Scat + WRN 12-8	**38.9 $\pm$ 1.2**	**54.7$\pm$0.6**	**62.0$\pm$1.1**

Single Shot MultiBox Detector (SSD)



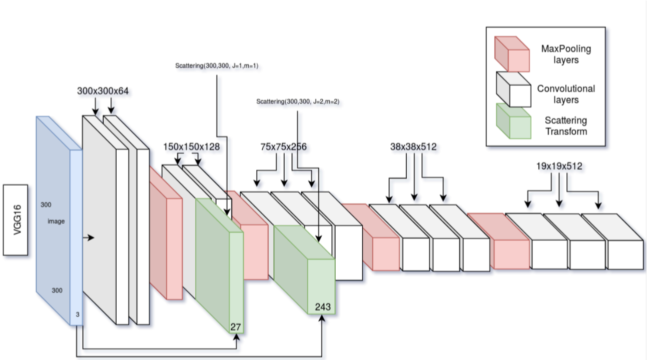
Sequential Scattering SSD

- Scattering is applied before data is piped through SSD



Parallel Scattering SSD

- Data is piped through scattering and standard SSD and continuously merged at different stages



Datasets - VOC

(A) (B) (C)

Figure 6: Three samples from the PASCAL VOC dataset showing a dog, bus and TV monitor from left to right.

Datasets - Kitti

(a)

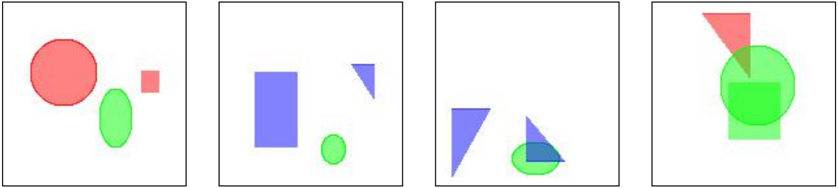
(b)

(c)

(d)

Figure 7: Four samples from the KITTI dataset.

Datasets - Toydata



(a) (b) (c) (d)

Figure 8: Four samples from the toy data set.

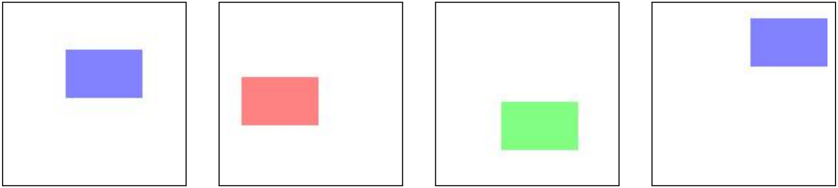
Toy datasets

Idea

Test if the promised equivariances/invariances hold on specifically created toy datasets

- Translation dataset
- Scale dataset
- Rotation dataset
- Deformation dataset

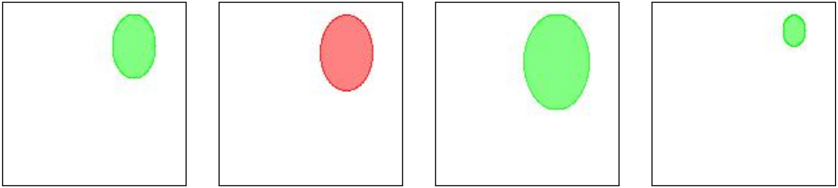
Translation dataset



(A) (B) (C) (D)

Figure 9: Four samples from the translation toy data set. a) is the base image; b) -d) are the translated versions

Scale dataset



(a) (b) (c) (d)

Figure 10: Four samples from the scale toy data set. a) is the base image; b) -d) are the scaled versions

Rotation dataset

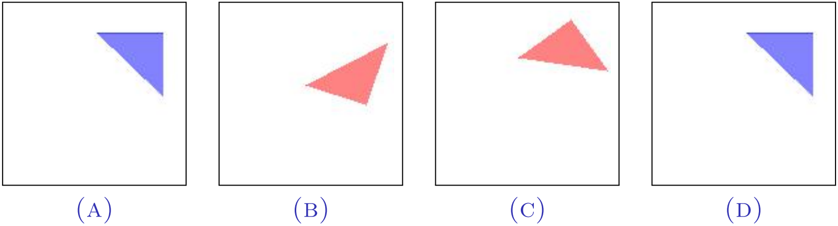


Figure 11: Four samples from the rotation toy data set. a) is the base image; b) -d) are the rotated versions

Deformation dataset

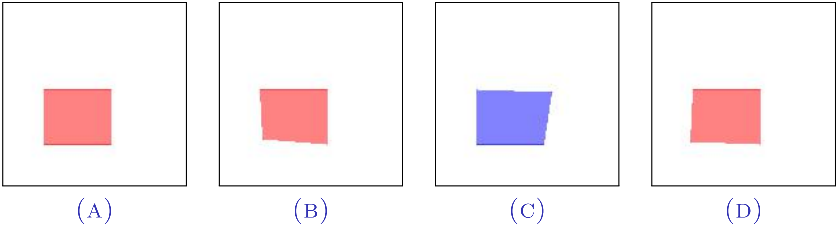


Figure 12: Four samples from the deformation toy data set. a) is the base image; b) -d) are the deformed versions

Experiments

1. Performance on all datasets
2. Performance on very small datasets with low training time
3. Time consumption per forward pass

Results - Comparison

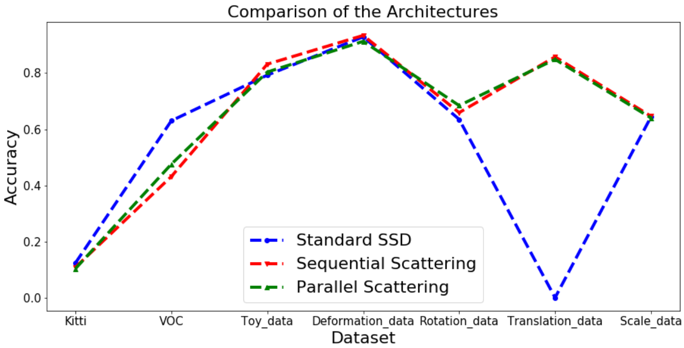


Figure 13: Final comparison on all datasets

Results - Small data experiments

Dataset	standard	sequential	parallel
Toy_data_small (25k)	0.630 ± 0.008	**0.759** ± **0.004**	0.411 ± 0.012
Toy_data_small (5k)	0.043 ± 0.007	**0.121** ± **0.027**	0.003 ± 0.001
VOC (25k)	**0.317** ± **0.011**	0.053 ± 0.006	0.013 ± 0.001
VOC (5k)	**0.025** ± **0.001**	0.011 ± 0.007	0.004 ± 0.000

Results - Timing experiments

network type	mean	std.
normal SSD	0.236	0.004
sequential scattering	0.178	0.004
parallel scattering	1.499	0.002

Conclusion

- The **sequential** Scattering Transform is faster and more robust method for some applications
- The **parallel** Scattering Transform is significantly slower and does not provide the supposed benefits
- (In a follow-up experiment the parallel scattering gets the best of both worlds while taking twice as long per forward pass)

Questions

Questions?

References

[BM12], [SM13], [OM14], [OBZ17], [ACC+17]

Tameem Adel, Taco Cohen, Matthan Caan, Max Welling, On behalf of the AGEhIV study group Initiative, and the Alzheimer's Disease Neuroimaging.
3d scattering transforms for disease classification in neuroimaging.
NeuroImage: Clinical, 14:506–517, 2017.
Exported from https://app.dimensions.ai on 2018/10/21.

Joan Bruna and Stéphane Mallat.
Invariant scattering convolution networks.
CoRR, abs/1203.1513, 2012.

Edouard Oyallon, Eugene Belilovsky, and Sergey Zagoruyko.
Scaling the scattering transform: Deep hybrid networks.
CoRR, abs/1703.08961, 2017.

Edouard Oyallon and Stéphane Mallat.
Deep roto-translation scattering for object classification.
CoRR, abs/1412.8659, 2014.

Laurent Sifre and Stephane Mallat.
Rotation, scaling and deformation invariant scattering for texture discrimination.
In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '13, pages 1233–1240, Washington, DC, USA, 2013. IEEE Computer Society.

Backup Equations - number of filters

$$i \cdot (1 + JL) \tag{2}$$

$$i \cdot (1 + JL + \frac{1}{2}J(J-1)L^2) \tag{3}$$

- Let $J = 2, L = 8, N, M = 32, 32$ for a RGB image.
- number of outputs of the scattering network for $m = 1$:

$$3 \cdot (1 + 2 * 8) = 51$$

- number of outputs of the scattering network for $m = 2$:

$$3 \cdot (1 + 16 + 0.5 * 2 * 1 * 64) = 243$$

- all outputs of size 8x8

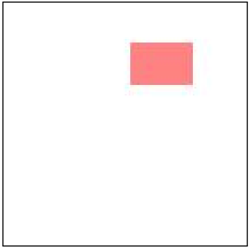
More definitions:

- Central Frequency:

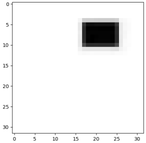
$$\int_{-\infty}^{\infty} \omega |\Psi(\omega)|^2 d\omega$$

where Ψ is the Fourier Transform of the wavelet ψ. This is the centre of mass of $|\Psi(\omega)|^2$.

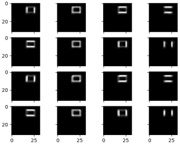
Coefficients - Example 1



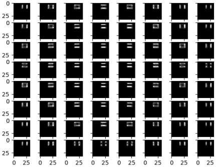
(A)



(B)

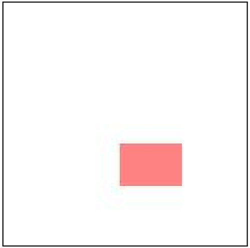


(C)

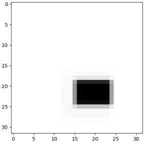


(D)

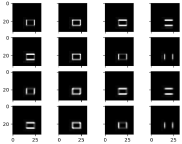
Coefficients - Example 1 (translated)



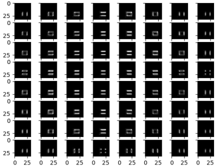
(e)



(f)

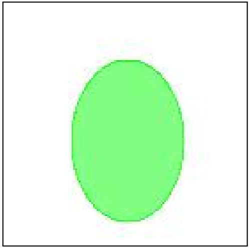


(g)

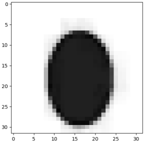


(h)

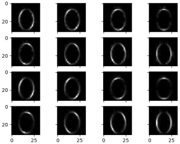
Coefficients - Ellipse



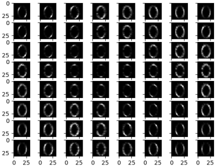
(i)



(j)

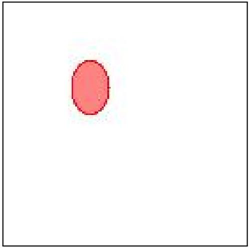


(k)

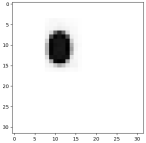


(l)

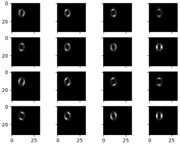
Coefficients - Ellipse (scaled)



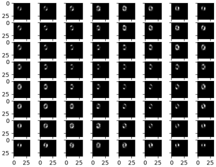
(m)



(n)



(o)



(p)