Deep Scattering Network with Max-pooling

Taekyung Ki[a] and **Youngmi Hur**[a,b]

[a]Yonsei University, Korea
[b]KIAS, Korea

DCC 2021

1. CNN and Scattering Network

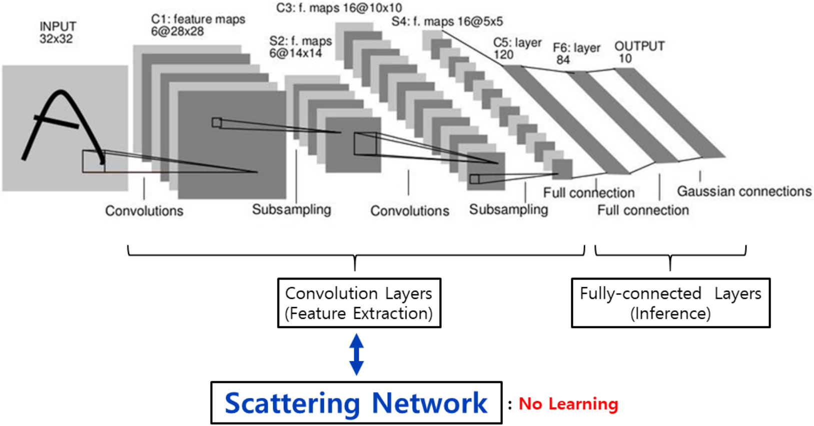

Figure 1: Two parts of CNN [1] and the role of scattering network [2].

Scattering Network

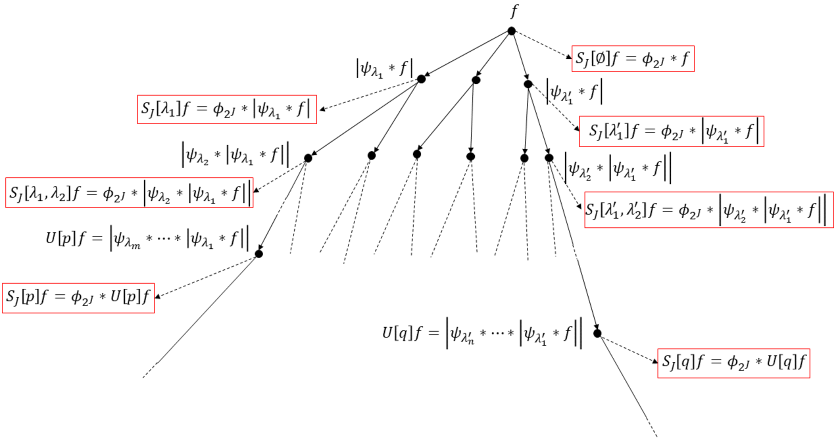

Figure 2: Propagation of scattering network. Input f is propagated toward deeper network by convolving with pre-defined wavelets followed by modulus. Two paths, $p = (\lambda_1, \cdots, \lambda_m)$ of length m and $q = (\lambda'_1, \cdots, \lambda'_n)$ of length n, are shown.

2. Scattering Network with Max-pooling [3]

Definition 1 (Continuous Plate)

Let $\mathcal{A}$ be a collection of $f : \mathbb{R}^d \to \mathbb{R}$ with compact support. If there exist rectangular compact regions $D, D^{(i)} \subset \mathbb{R}^d$, $i = 1, 2, \cdots, N$ satisfying

- $\bigcup_{f \in \mathcal{A}} \mathrm{supp}(f) \subset D$ and $0 \in D$,
- $D = \bigcup_{i=1}^{N} D^{(i)}$,
- $\forall i = 1, 2, \cdots, N,\ |D^{(i)}| = |D|/N$,
- $\exists k \in \mathbb{R}^d$ such that $\left|\left\{x \in \mathbb{R}^d : x \in \left(D^{(i)} - (D^{(j)} + k)\right)\right\}\right| = 0$ if $i \neq j$,

we say D is a **continuous plate** of $\mathcal{A}$ and $D^{(i)}$ is the i-th sub-plate of D.

Definition 2 (Continuous Max-pooling)

Let $\mathcal{A}$, D and $D^{(i)}$ be defined as in Def. 1. Let $L^\infty(D) := L^\infty(\mathbb{R}^d) \cap \mathcal{A}$. For $f \in L^\infty(D)$ and $S > (|D| \|f\|_\infty / \|f\|_2)^{1/d}$, a **continuous max-pooling operator** $P : L^\infty(D) \to L^\infty(D)$ (with a pooling factor S) is defined by

$$P(f)(x) := \sum_{i=1}^{N} \|f \chi_{D^{(i)}}\|_\infty \chi_{D^{(i)}}(Sx).$$

(Note: $|D| \|f\|_\infty / \|f\|_2 \geq 1$, hence $S > 1$.)

Definition 3 (Windowed Pooled-Scattering Transform)

Let $J \in \mathbb{Z}$ be fixed, and let $\lambda := (j, r)$, $j > -J$, $r \in G^+$.
For pre-defined wavelet $\psi \in L^2(\mathbb{R}^d)$, let $\psi_\lambda(x) := 2^{dj}\psi(2^j r^{-1}x)$.
Let P be the continuous max-pooling with a pooling factor S.
Let $q = (\lambda_1, \lambda_2, \cdots, \lambda_m)$ be a path of length m. We define

- $\tilde{U}[\emptyset] = P$
- $\tilde{U}[\lambda] : L^\infty(D) \to L^\infty(D/S) : f \mapsto \tilde{U}[\lambda]f = P(|\psi_\lambda * f|)$
- $\tilde{U}[q] : L^\infty(D) \to L^\infty(D/S^m) : f \mapsto \tilde{U}[q]f := \tilde{U}[\lambda_m]\tilde{U}[\lambda_{m-1}]\cdots\tilde{U}[\lambda_1]f$

A **windowed pooled-scattering transform** $\tilde{S}_J[q]$ is defined by

$$\tilde{S}_J[q]f := (\phi_{2^J} * \tilde{U}[q]f), \quad \forall f \in L^\infty(D),$$

where $\phi_{2^J}(x) := 2^{-dJ}\phi(2^{-J}x)$ is a low-pass kernel from $\phi \in L^2(\mathbb{R}^d)$.

Scattering-maxp Network

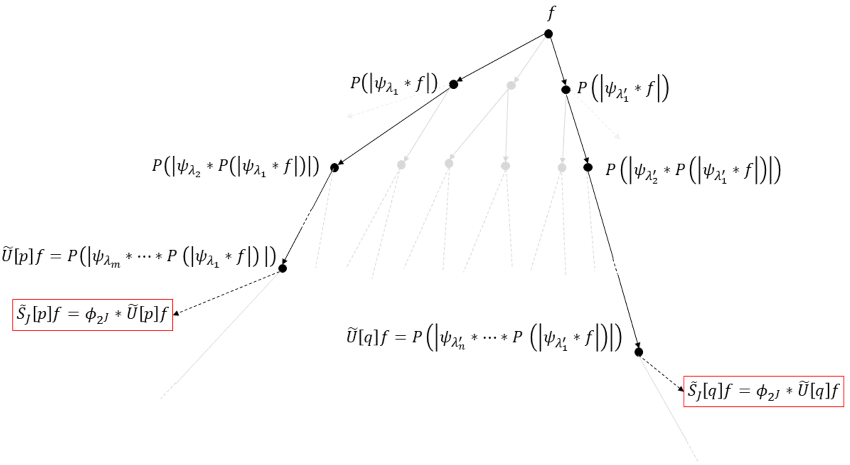

Figure 3: Propagation of **scattering-maxp network**. It is obtained by the windowed pooled-scattering transform. Two paths, p and q, are highlighted.

Translation Invariance of Scattering-maxp Network

Theorem 1. (Translation Invariance)

Suppose $c \in \mathbb{R}^d$ satisfies $0 \in D + c$ and $|\hat{\phi}(\omega)|(\omega)| < B$ a.e. $\omega \in \mathbb{R}^d$ for some $B > 0$. Then

$$\lim_{m\to\infty} \sum_{p\in\Lambda_J^m} \|\tilde{S}_J[p]f - \tilde{S}_J[p]T_c f\|_2^2 = 0, \quad f \in L^\infty(D)$$

where Λ_J^m is the set of all possible paths $p = (\lambda_1, \cdots, \lambda_m)$ of length m, $\hat{\phi}$ is the Fourier transform of ϕ, and $T_c f(x) = f(x - c)$ is the translation by c.

3. Experimental Results [3], [4]

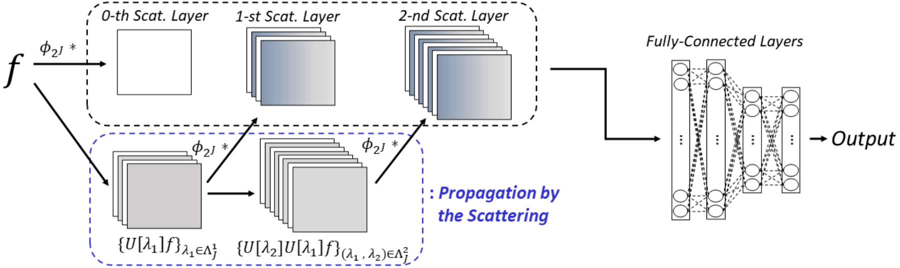

Figure 4: **Scattering**. f is an input. 4 fully-connected layers are added right after computing the output of scattering. The output of the model is calculated by soft-max.

Scattering-maxp

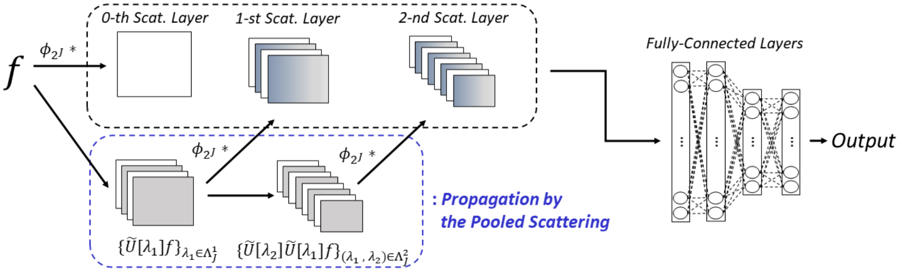

Figure 5: **Scattering-maxp**. It is based on the original one as in Figure 4 but two 2×2 max-poolings (with 2×2 stride) are inserted right after the scattering layer-1 and layer-2.

Scattering-naivep

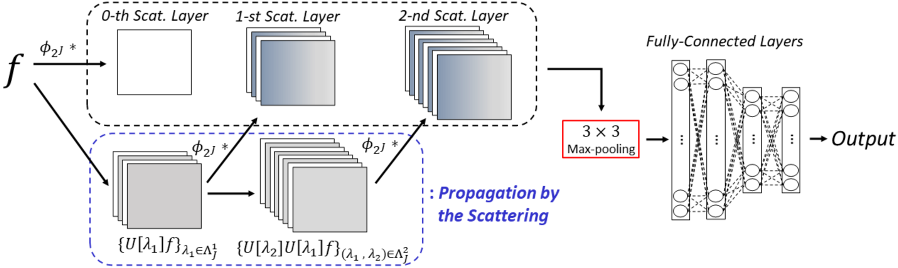

Figure 6: **Scattering-naivep**. It is based on the original one as in Figure 4 but additional uses 3×3 max-pooling (with 3×3 stride) right after calculating the output of scattering.

Training Set-up

- CPU: Intel(R) Xeon(R) Gold 5210 @ 2.20GHz / GPU: Tesla V100-32GB.
- TensorFlow 1.15.1 / Keras 2.2.4-tf.
- Loss: Categorical cross-entropy / Optimizer: Adam
- Data Augmentations: Horizontal flip & 8 rotations from -20° to 20°.
- Batch Size: 256
- Ratio of the Training Data and the Validation Data: 3 to 1.
- Input Size: 224×224.

Datasets: Caltech-101 / Caltech-256

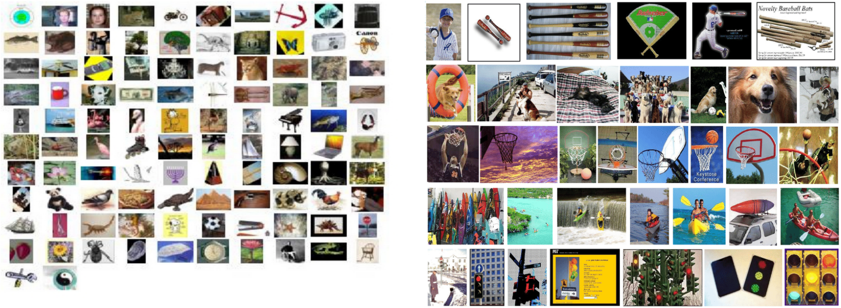

Figure 7: Datasets for experiments: (Left) Caltech-101 [5] (Right) Caltech-256 [6]

Experimental Results on Caltech-101

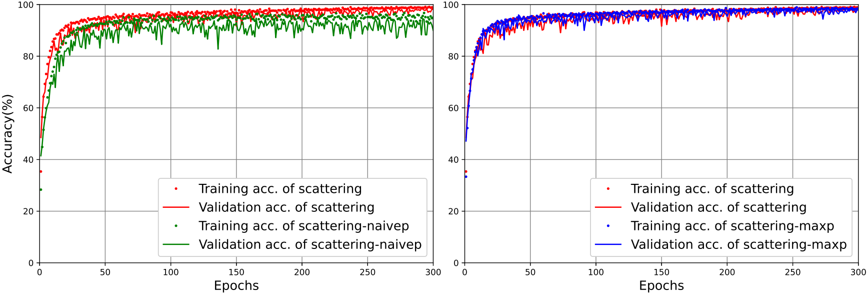

Figure 8: Performance of scattering based models on Caltech-101 (300 epochs). (Left) Original scattering and scattering-naivep. (Right) Original scattering and scattering-maxp. The learning rate is 10^{-3} during the training session.

Experimental Results on Caltech-101

Table 1: Classification results on Caltech-101.

Model	# of parameters	Accuracy (%)	Training time (s/epoch)
VGG-16	134,677,286	99.58	566
ResNet34	21,344,166	99.21	221
MobileNet	**3,332,742**	**99.95**	419
Scattering	87,592,038	98.49	284
Scattering-naivep	11,596,902	94.54	267
Scattering-maxp	9,944,166	98.59	**206**

Experimental Results on Caltech-256

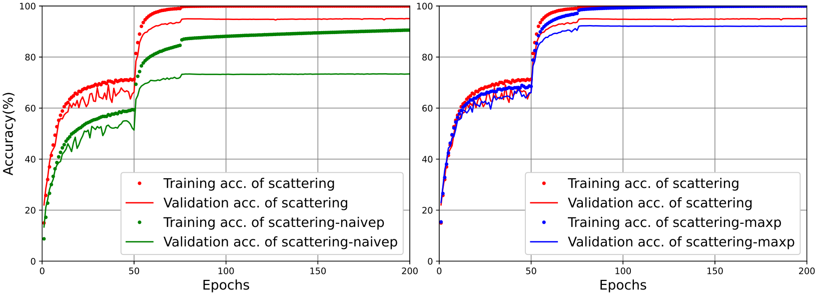

Figure 9: Performance of scattering based models on Caltech-256 (200 epochs). (Left) Original scattering and scattering-naivep. (Right) Original scattering and scattering-maxp. The learning rate is adjusted to be 10^{-3} from the beginning, 10^{-4} after 50 epochs, and 10^{-5} after 75 epochs.

Experimental Results on Caltech-256

Table 2: Classification results on Caltech-256.

Model	# of parameters	Accuracy (%)	Training time (s/epoch)
VGG-16	135,312,321	98.58	1926
ResNet34	21,423,681	99.97	750
MobileNet	**3,491,617**	**99.99**	1433
Scattering	87,631,873	95.06	1006
Scattering-naivep	11,636,737	73.38	982
Scattering-maxp	9,984,001	92.11	**651**

References

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, *Backpropagation Applied to Handwritten Zip Code Recognition*, Neural Computation, vol.1, pp.541-551, 1989

[2] S. Mallat, *Group Invariant Scattering*, Communications on Pure and Applied Mathematics, vol.65, pp.1331-1398, 2012.

[3] T. Ki and Y. Hur, *Deep scattering network with Max-pooling*, preprint, (https://arxiv.org/abs/2101.02321), 2021.

[4] https://github.com/TaekyungKi/Scattering_maxp.

[5] L. Fei-Fei, R. Fergus and P. Perona., *Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories*, in Conference on Computer Vision and Pattern Recognition Workshop, 2004.

[6] G. Griffin, A. Holub, and P. Perona., *Caltech-256 object category dataset*, preprint, 2007.