Exposé: Object Detection using the Scattering Transform

1 Outline

1. **What does already exist?** There is a paper introducing the scattering transform [BM12]. There are multiple papers showing that a combination of Deep Learning architectures and the Scattering Transform can yield state of the art (at their point in time) results for classification tasks and could outperform classic Deep Learning methods for small datasets with some deformations [SM13] [OBZ17] [ACC+17].

2. **What is new?** There are two new contributions: a) The Scattering Transform is applied to object detection with the same method used for classification. This has not been done before (which I will call sequential scattering). b) A second technique is introduced which combines the Scattering Transform and Deep Learning methods in a different way (which I will call parallel scattering).

3. **What are the insights/results?** There are two main results:
   a) The sequential scattering is outperformed by conventional methods on some datasets and outperforms conventional methods on others. It also has some theoretical guarantees that conventional methods do not have. It also 25% faster than the conventional method. Additionally, it needs less samples to generalize on some datasets compared to conventional methods. In my opinion it is reasonable to use the sequential scattering when you want to have specific theoretical guarantees and speed is important.
   b) The parallel scattering gets the best of both worlds. It is as good as the best of conventional methods and sequential scattering in tested cases (Note: this is a result from a follow up experiment and not from the finished bachelors thesis). It also provides the theoretical guarantees that the sequential scattering has. The downside of the parallel approach is that it takes around twice as long to compute a forward pass as the conventional method. In my opinion the parallel approach can be used in cases where a robust model is more important than fast training.

4. **Why are the results relevant?**
   - Providing theoretical guarantees for some kind of transformations can be made without loosing much accuracy. This is be very important in some tasks, i.e. handwritten digit recognition.
   - Generalizing patterns better than conventional methods from small amounts of samples can be useful for some applications, i.e. medical applications where only few samples are available.
   - Making a forward pass 25% faster can be important for some online applications like autonomous driving where fast computations are a necessary condition for a working system.

5. **Why do the new methods work better on the benchmarks** There are three questions which must be answered to understand the results of this method.
   - Why are the results of the scattering methods and the conventional network almost equally good on nearly all datasets? Mainly because all the important information for 2D object detection is contained in a representation that focuses on edges. The scattering representation is a sufficient representation for object detection.
   - Why are the results of the scattering methods better on the translation dataset? Because the scattering transform provides the guarantee of local equivariance w.r.t. translations while the conventional SSD has nearly no theoretical guarantees.
   - Why are the results of the sequential scattering worse than the conventional SSD on VOC? Probably because VOC has overlapping objects in some of its pictures. Two objects can therefore not be perfectly reconstructed only through their edge information if they overlap too much.

2 Questions

1. Do we need state of the art algorithms? We use a VGG as the object detection network. VGG is a fully convolutional detector and therefore faster than the two-stage detectors like Faster-RCNN. Therefore we cannot show that our methods beat two stage detectors on the benchmark sets. However, I am not sure if that is a necessary condition for the results to be of importance to the scientific community. We show that the small additions to a VGG setup (sequential and parallel scattering) have specific advantages compared to a conventional VGG setup (without the additions). I feel like this already is an important addition to scientific progress and using two stage detectors is only a nice to have but not necessary addition (We also argue why the method is easily extendable to two stage detectors). However, you have significantly more experience so I would like to hear your opinion on this issue.

Literatur

[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.

[BM12] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *CoRR*, abs/1203.1513, 2012.

[OBZ17] Edouard Oyallon, Eugene Belilovsky, and Sergey Zagoruyko. Scaling the scattering transform: Deep hybrid networks. *CoRR*, abs/1703.08961, 2017.

[SM13] Laurent Sifre and Stéphane Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1233–1240, 2013.