Parent Item: Neural Nearest Neighbors Networks

Summary

1 Intro

- N^3 = neural nearest neighbors block
- demonstrate that they can improve non-local processing by optimizing feature space for matching
- Make three claims:
    1. propose continuous deterministic relaxation of KNN rule
    2. N^3 block still enables end-to-end trainable non-local processing
    3. accuracy of 2 image tasks improve significantly by adding N^3 blocks to strong local CNN architectures

2 Related Work

- There are previous works on weighted avgs being used for building differentiable kNN classifiers, this work proposes differentiable kNN selection rule (to output a set of neighbors instead of a single aggregation of labels) (13, 35, 41)
- Their other difference is that they are making a differentiable KNN selection rule to keep the end-to-end functionality

3 Differentiable k-Nearest Neighbors

- First they interpret deterministic KNN rule as limit of a parametric family of discrete stochastic sampling processes then derive continuous relaxations for the discrete vars
- Consider a continuous deterministic relaxation because gradient estimators for expectations over discrete vars suffer from high variance
- Want to replace one-hot coded weight vectors with continuous expectations
- The relaxation still contains hard KNN selection rule as the limit case
- You are able to train temp t since the NN are differentiable with respect to it

4 Neural Nearest Neighbors Block

- Use concept of self-similarity (query set and database are derived from same features)
- Steps:
    1. embedding network takes input to produce feature embedding and temp params
    2. Those are used to compute continuous nearest neighbors feat volumes that are aggregated with the input
- These blocks are interleaved with common networks to improve the overall results from the architecture

5 Experiments

- Use image denoising as main test for how effective N^3 is
- Show that their new design outperforms local methods AND competing non local models

6 Conclusion

- Created a new block that can be inserted in other Neural Nets and can outperform current techniques
- Expect this block to also benefit end-to-end processes for other input domains (for sequence-valued data)