Neural Nearest Neighbour Network

Project ID : 15 | Robot Chitti

Kartik Mehta (2019102029)
Abhigyan Gargav (2019102040)
Rohan Lahane (2019112004)
Raj Singh Parihar (2019102028)

GitHub Link:

Problem Statement

Convolutional neural networks (CNNs) have revolutionized many areas of machine learning by providing great predictive accuracy. They mainly focus on local processing by combining convolutional layers and element-wise operations. So there is a trade off between context size and localization accuracy.
There are many traditional non-local methods for image restoration and denoising. They use the self-similarity in images. Recently, this type of non-local processing is being incorporated in Neural Networks. The existing approaches for this rely on KNN (k-nearest neighbour), but the main issue in them is the non-differentiability of the KNN selection rule.

The problem is that we need to come up with a better (more optimizable) way of feature classification than KNN. Since KNN cannot be optimized more using traditional methods, this paper helps us understand how KNN is represented as a differentiable mathematical model, so that it can be optimized using traditional methods.

After finding this optimized model, we plan to implement it to solve real world problems like gaussian image denoising and compare the results of our implementation with a bunch of other models. E.g. already implemented state of the art models to solve which are problem specific.

So, The **first part** of our project is:

We need to come up with an approach to find a relaxation of the KNN rule, which allows differentiation. Then using the relaxation, develop a novel neural network layer which enables end-to-end trainable non-local processing based on the principle of self-similarity.

And the **second part** of our project is:
To show that this method and architecture outperforms strong non-local approaches that rely on KNN and CNN.

Goals And Approach

The main goals and their approaches are:

1. Proposing a relaxation for KNN rule so that it allows differentiability of the output w.r.t pairwise distances of the inputs.

   The KNN selection rule is defined as follows:
   Assume that we are given a query item q, a database of candidate items $(x_i)\ i\ \in I$ with indices $I = \{1, 2, \cdots, M\}$for matching, and a distance metric $d(.,.)$ between pairs of items.. Let $\pi_q : I \rightarrow I$be a permutation that sorts the database items by increasing distance to q:

   $$\pi_q(i_1) < \pi_q(i_2)$$

   This implies, $$d\ (q, x_{i_1}) < d(q, x_{i_2})$$

   Where, $$i_1, i_2 \in I$$

   The KNN of q are then given by the set of the first $k$items w.r.t. the permutationπ_q

   $$KNN(q) = \{x_i \,|\, \pi_q(i) \leq k\}$$

   This rule is not differentiable. This does not allow us to derive gradients. So we will first show KNN as a limit of a parametric family of discrete stochastic sampling processes. Then we will derive continuous relaxations for the discrete variables. So we would eventually end up with a continuous deterministic relaxation.

   KNN rule as limit distribution
   KNN is interpreted as the limit of k categorical distributions. Let $Cat(w_1 | \alpha_1, t)$be a categorical distribution, such that:

   $$Cat(w_1 | \alpha_1, t) = P[w_1 = i | \alpha_1, t]$$

   Where, $\alpha_1 =- d(q, x_i)$ and w= all zero vectors with 1 at i^{th} position.

   In the limit $t \rightarrow 0$, $Cat(w_1 | \alpha_1, t)$ will converge to a deterministic distribution centered at the index of the database item with the smallest distance to q. Hence $Cat(w_1 | \alpha_1, t)$ is a stochastic relaxation of 1-NN. We need to generalise it for K-NN by constructing further distributions $Cat(w_{j+1} | \alpha_{j+1}, t)$
   Having derived KNN as a limit of a parametric family of discrete stochastic sampling processes, we will now derive continuous relaxations for the discrete variables.

Continuous deterministic relaxation

We will replace the weighted vectors w with their continuous expectations. This will give us a continuous relaxation of the stochastic nearest neighbors that still converges to the hard KNN selection rule in the limit case of $t \rightarrow 0$.

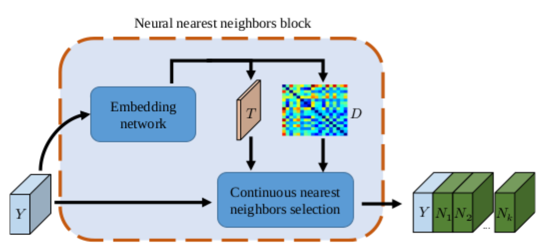
The expectation $\bar{w}_1$ of w_1 is given by:

$$\bar{w}_1 = P[w_1 = i \,|\alpha_1, t]$$

2. <u>Develop a neural network layer which enables end-to-end trainable non-local processing based on the principle of self-similarity.</u>

N^3 - block consists of two important parts.

First, As shown in the figure, an embedding network takes the output Y of the previous layer as input and calculates a pairwise distance matrix D between elements in Y as well as a temperature parameter(T)



(a) N^3 block

The pairwise distance matrix D can be obtained by $D_{ij} = d(E_i, E_j)$ where $E_i = f_E(Y)$.
Is the calculated feature embedding for ith element , and d is a differentiable distance function Another network branch computes a tensor $T = f_t(Y)$ containing the temperature 't' for each item.

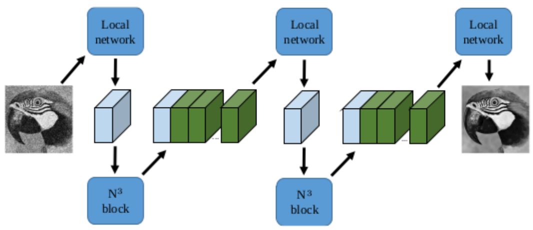
In the **second step** we use these to compute a stack of continuous nearest neighbors feature volumes that are concatenated with the input(Y). We interleave blocks with existing local processing networks to form neural nearest neighbors networks (N^3 Net) Y and each N i have equal dimensionality, we can use an element-wise operation to aggregate the original features Y and the neighbors. Thus we concatenate Y and the N_i along the feature dimension, which allows further network layers to learn
how to fuse the information effectively in a non-linear way

3. <u>Demonstrate that the accuracy of image denoising can be improved significantly by augmenting strong local CNN architectures with our novel N^3 block and compare it with non-local baselines.</u>

   Defining Architecture

   We will augment the architecture of the DnCNN model with our N^3 net for image denoising.The combined architecture is shown below:

   Local network : DnCNN



(b) N^3Net

   Comparison with state of the art

   We compare the N^3 net with various models such as DnCNN baseline, BM3D, NLNet etc.

Expected Deliverables :-

From the implementation of the project, the expected deliverables are -

- An efficient implementation of Neural Nearest Neighbour block (N^3 Block)
- Improving non-local processing by optimizing the feature space for matching.
- Trying different non-local and other methods like BM3D Method along with N^3 Block for comparison.
- To demonstrate that N^3 Block outperforms strong convolutional neural network (CNN) baselines and recent non-local models that rely on KNN selection in hand-chosen features spaces.
- We propose a differentiable relaxation of the KNN selection rule where the output is a set of neighbors, instead of a single aggregation of the labels of the neighbors.
- Presentation and description of the project along with instructions to run the demo.

Timeline & Milestones :-

Timeline	Milestones
26 Oct. 2021	Project Allocation
27 - 6 Nov. 2021	Paper and relevant work reading for Proposal.Discussed doubts with TA.
7 Nov. 2021	Project Proposal Submission
7 - 17 Nov. 2021	Implementing Neural Nearest Neighbour block (N^3 Block) and Basic understanding about different methods.
17 - 20 Nov. 2021	Mid-Evaluation
21 Nov - 1 Dec 2021	Running the method on proposed application and presentation preparation. Finishing steps and Final Submission
1 - 4 Dec. 2021	Final Presentation
4 Dec. 2021	Final Submission

Work Distribution :-

Part 1 *(as per problem statement)*

Abhigyan Gargav

Rohan Lahane

Part 2 *(as per problem statement)*

Kartik Mehta

Raj Singh Parihar

References :-

1.Stamatios Lefkimmiatis. Universal denoising networks: A novel CNN-based network architecture for image denoising. In CVPR, pages 3204–3213, 2018.

2.Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In CVPR, pages 60–65, 2005.

3.https://www.youtube.com/watch?v=YRhxdVk_sIs - For Basic Understanding of CNN in image pattern recognition
.