Problem Definition

Goal: Few-shot Learning (FSL) which aims to learn from few labeled training data and train a good classification model with few labeled data.

Definition:
For few-shot classification, the dataset contains a meta-train set and a meta-test set, which have disjoint categories (Cmeta-train ∩ Cmeta-test = ∅). The meta-train set, in which each category contains many labeled samples, is used to train a generalized model. Then, it is evaluated on the meta-test set with only few labeled images in each category. In the popular setting, many testing episodes are performed during evaluation process. Each testing episode contains N categories and each category has K labeled images. These labeled data form the support set S = {(x1, y1), (x2, y2), ..., (xN, yN)}, where (xi, yi) is the datum of ith class. The rest unlabeled testing data of these N categories form the query set Q = {(x1, y1), (x2, y2), ..., (xN, yN)}. Such setting is called N-way K-shot learning.

Motivation:
- Previous FSL method learns to compare the testing (query) image and training (support) image by simply concatenating the features of two images and feeding it into the neural network.
- With few labeled data in each class, the neural network has difficulty in learning or comparing the local features of two images. Such simple image-level comparison may cause different kinds of misclassification.
- For more precise comparison, we propose Augmented Bi-path Network (ABNet) to learn to compare both global and local features on multi-scales.

Method
To fully exploit the few-labeled images, we propose Augmented Bi-path Network (ABNet) for Few-shot Learning, the framework is illustrated as follow.

Network Architecture:

Algorithm:
1. **Salient Patch Extraction:** starts with the patches sampled by selective search method and selects top5 Salient Patches measured by:
   \[
   S(c_i) = \frac{1}{N} \sum_{j=1}^{N} f(c_i, y_j) - \frac{1}{N} \sum_{j=1}^{N} f(c_i, y_j)
   \]
2. **Feature Embedding:** encode the whole image and salient patches by CNN backbone f.
3. **Learning to Augment:**
   \[
   \left(\begin{array}{c} x_i \\ y_i \\ z_i \end{array}\right) = A \left(\begin{array}{c} x_i \\ y_i \\ z_i \end{array}\right) = \left[\begin{array}{ccc} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{array}\right] \left(\begin{array}{c} x_i \\ y_i \\ z_i \end{array}\right)
   \]
4. **Learning to Compare:**
   a) Generating Similarity Maps
   b) Learning to Re-weight
   c) Learning to Merge
   \[
   \alpha(x, q) = \left\{ \begin{array}{ll}
   \frac{1}{N} & \text{if } y_i = y_j \\
   0 & \text{otherwise}
   \end{array} \right.
   \]

Loss Function:
- **Classification Loss:**
  \[
  \mathcal{L}_{cls} = \frac{1}{B} \sum_{b=1}^{B} \sum_{k=1}^{K} f(c_b, y_k) - \mathbb{E}[f(c_b, y_k)]^2
  \]
- **Attention Loss:**
  \[
  \mathcal{L}_{att} = \frac{S(a_{i,j})}{N} \sum_{k=1}^{K} \mathbb{E}[f(c_b, y_k)]
  \]
- **Augmentation Loss:**
  \[
  \mathcal{L}_{aug} = \frac{S(a_{i,j})}{N} \sum_{k=1}^{K} \mathbb{E}[f(c_b, y_k)]
  \]
- **Total Loss:**
  \[
  L = \mathcal{L}_{cls} + \lambda_{att} \mathcal{L}_{att} + \lambda_{aug} \mathcal{L}_{aug}
  \]

Ablation Studies
Quantitative comparison of four variants of our methods:
- **Baseline:** fixed global features
- **Baseline + LA (Learn to Augment)**
- **Baseline + LA + SP (Salient Patch)**
- **Baseline + LA + SP + LR (Learn to Re-weight)**

Visualization Comparison:

Results:

**Ablation Studies**

**Visualization Comparison**

Conclusion
We propose a novel meta-learning based method, namely Augmented Bi-path Network, for Few-shot Learning. The proposed method extends the previous “learn-to-compare” based methods by introducing both global and local features on multi-scales, and learns to augment the features for better robustness. Experimental results show that our method significantly outperforms the state-of-the-art on three challenging datasets under all settings.