A CONTRIBUTIONS

• Reformulating convolution block based local feature embedding as feature assignment through best matching kernel
• Repurposing softmax as a batch-statistics-free replacement of BN-ReLU
• Exploiting mixture of class labels to shape the intermediate features of CNN

AN APPROACH

Low level features    Mid level features    High level features

CNNs: Extracting hierarchical features through stacked convolution blocks whose parameters are learned in top-down manner via feedback from a class-supervised loss function.

• Differently, we want to use the mixture of class labels to synthesize new labels that can explicitly supervise lower level feature extraction (e.g. plane + bird ≈ wing).
• We first formulate an optimization problem to analytically express selecting the best matching kernel and assigning its semantic vector.

THEORY TO PRACTICE

• Given a set of matching kernels \( \omega_k \) and 3x3 features as \( x_{3 \times 3} \), we define the problem as:
\[
p^* = \arg \max_{q \in [0,1]} q + \sum_k \frac{1}{\sum_k} \omega_k^T x_{3 \times 3}
\]
• We make the above operation differentiable by smoothing the objective with entropy.
• We represent 3x3 pattern with \( x' = \sum_k p_k^* \omega_k \).
• Relating BN-ReLU to softmax, we show that 3x3-BN-ReLU-1x1 inherently perform these operations.

IMPLICATIONS

• We observe better clustering effects in which a clear distinction between animals and vehicles exists.
• Our method enables mixture of class labels and creates fine labels by using existing classes with the help of our embedding vectors which semantically reshape the overall geometry.
• Our formulation mimicks cross-attention between \( x_{3 \times 3} \) (queries) and \( \omega_k \) (keys) in which the final representation is given as the convex combination of \( \omega_k \) (values).

METHOD OVERVIEW

1. We augment the training loss with an auxiliary per patch \( (x_{\square}) \) classification loss and enable label supervision in lower layers:
\[
L(D) = \frac{1}{|D|} \sum_{(I,y) \in D} \left( 1 - \lambda \right) f(h_1(I), y) + \frac{1}{\lambda} \sum_{x \in h_1(I)} \lambda L(h_2(x_{\square}), y)
\]
2. With our new block, we yield novel semantic entities as the convex combination of the class features and explicitly shape the feature embeddings according to these semantics.

ANALYSIS OF BEHAVIOUR

• We generate new semantic entities from the combination of class vectors such as wing from bird and plane, tire from car and truck.
• We observe discriminative patches specific to certain classes.
• We observe generic entities as the mixture of many classes such as fur for animal classes.

ABLATION STUDIES

(a) BN-ReLU \( \approx \) soft-maximizer
(b) The inclusion of our mechanism through auxiliary loss boosts the performance.

QUANTITATIVE RESULTS

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Cifar10</th>
<th>Cifar100</th>
<th>Mini-ImageNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNResNet</td>
<td>0.96M</td>
<td>50.92</td>
<td>65.94</td>
</tr>
<tr>
<td>BNResNet+aux.</td>
<td>0.96M</td>
<td>50.97</td>
<td>66.21</td>
</tr>
<tr>
<td>BNResNet+ours.</td>
<td>0.96M</td>
<td>51.6C</td>
<td>66.78</td>
</tr>
<tr>
<td>RN26</td>
<td>1.42M</td>
<td>90.78</td>
<td>68.35</td>
</tr>
<tr>
<td>RN26+ours.</td>
<td>1.44M</td>
<td>91.56</td>
<td>69.01</td>
</tr>
<tr>
<td>WRN16</td>
<td>1.29M</td>
<td>89.52</td>
<td>67.11</td>
</tr>
<tr>
<td>WRN16+ours.</td>
<td>1.30M</td>
<td>91.10</td>
<td>67.56</td>
</tr>
<tr>
<td>DNN100</td>
<td>1.25M</td>
<td>92.62</td>
<td>71.25</td>
</tr>
<tr>
<td>DNN100+ours.</td>
<td>1.32M</td>
<td>92.92</td>
<td>71.25</td>
</tr>
<tr>
<td>DNN100+ours.</td>
<td>1.36M</td>
<td>92.71</td>
<td>72.14</td>
</tr>
</tbody>
</table>