

## 1. はじめに

深層学習のホワイトボックス化として、認識時に判断根拠を可視化し視覚的説明を行うことは重要である。Attention Branch Network(ABN)[1] は、視覚的説明だけではなく、判断根拠をアテンションとしてアテンション機構に応用することで可視化に貢献している。一方で、物体認識では学習データとそのバリエーションを確保するために画像合成して水増しするデータ拡張が活用されている。特に画像を合成するアプローチは精度向上に大きく寄与している。しかしながら ABN において、画像合成を行うデータ拡張の影響が十分に考慮されていない。そこで、本研究では画像合成を行うデータ拡張を追加することで、ABN の判断根拠および精度に対してどのような影響が出るのかを調査する。

## 2. Attention Branch Network(ABN)

ABN は、物体認識時の判断根拠となる注目領域をアテンションマップとして可視化しつつ、認識に利用する手法である。図 1 のように ABN は、Feature extractor, Attention branch, Perception branch の 3 つのモジュールから構成されている。Feature extractor は、入力画像に対する特徴マップを抽出する。抽出した特徴マップを Attention branch に与え、アテンションマップを獲得する。アテンションマップは入力画像に対する注視領域を示し、Feature extractor の特徴マップに乗算し、Perception branch に与える。Perception branch は認識結果を出力する。

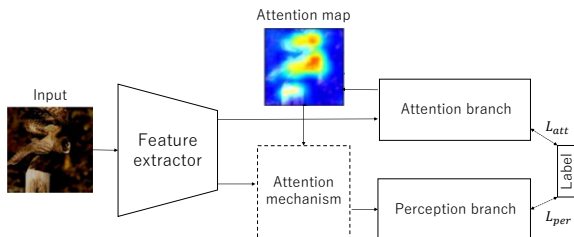


図 1: Attention Branch Network の構成

## 3. データ拡張

データ拡張には、平行移動・拡大縮小などの幾何学変換を加える方法、明るさやノイズを加える方法、画像を合成する方法がある。画像を合成する方法には、2 つの画像とラベルを混合する mixup、画像内のある領域をランダムに黒く塗りつぶす cutout、画像の一部にランダムな画像をはめ込む cutmix がある。合成したデータを学習用データに追加することで過学習を抑制でき、性能の向上が期待できる。図 2 に各手法の例を示す。



(a) mixup (b) cutout (c) cutmix

図 2: データ拡張の例

## 4. 調査

ABN は、アテンションマップを特徴マップに反映させることで認識に重要となる領域に注目した認識結果となる。そのため、データ拡張により合成された画像を用いると、アテンションマップに悪影響を及ぼすことが考えられる。本研究では、mixup, cutout, cutmix の 3 つのデータ拡張手法を ABN の学習に用いて、それらの影響を調査する。

### 4.1. 実験条件

本実験では、データセットに CIFAR100 と ImageNet を用いる。学習パラメータはバッチサイズを 128, epoch 数を CIFAR100 では 300, ImageNet では 200 とする。学

習係数を 0.1 に設定し、CIFAR100 では 150 と 225epoch, ImageNet では 100 と 150epoch の時に学習係数を 0.1 倍する。Top1 精度に加え、判断根拠の定量的評価には insertion と deletion を用いる。insertion は重要な領域を追加していき、その時の認識精度がわずかな領域でも向上すると良い。deletion は重要な領域を削除していき、その時の認識精度が領域でも低下すると良い。これらの領域を追加または削除する割合を変え、その時の Area Under Curve (AUC) で評価する。

## 4.2. 実験結果

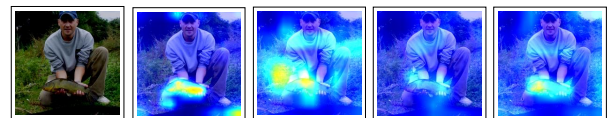
CIFAR100 と ImageNet での評価結果を表 1 に示す。CIFAR100 では、ResNet20 と Resnet152 どちらもデータ拡張を用いることで精度が向上していることが確認できる。また、ネットワークモデルによって最も性能の良いデータ拡張手法が異なることがわかった。ImageNet では、ResNet18 において cutout の精度向上が見られたが、mixup と cutmix では精度の向上が見られなかった。ResNet152 においては、どのデータ拡張を用いても精度が向上していることが確認できた。これらの結果から、大規模なデータセットを対象とする場合ネットワークサイズの小さいモデルに対しては、画像データのラベルを変えない cutout が有効であり、ネットワークサイズの大きいモデルに対しては、2 枚の画像を合成するような mixup や cutmix が有効であると考えられる。

表 1: 実験結果

データセット	CIFAR100		ImageNet	
	ResNet20	ResNet152	ResNet18	ResNet152
DA なし	69.32	76.78	70.34	77.87
+mixup	70.19	79.71	67.64	77.90
+cutout	<b>72.58</b>	78.79	<b>70.87</b>	78.74
+cutmix	71.05	<b>80.80</b>	68.07	<b>79.32</b>

ImageNet でのアテンションマップの可視化結果を図 3 に示す。図 3 より、データ拡張を用いることで精度は向上したが、mixup と cutmix では注視する領域が増えている。一方、cutout では注視領域が限定されている。

アテンションマップの評価結果を表 2 に示す。表 2 より、データ拡張前と比較して精度向上は見られない。これより、mixup ベースのデータ拡張は、精度向上に貢献するが、アテンションマップに悪影響を及ぼすことが分かった。



(a) 画像 (b) ResNet152(c)+mixup (d)+cutout (e)+cutmix

図 3: アテンションマップの比較

表 2: insertion と deletion による定量的評価 (AUC)

ネットワーク	insertion ↑	deletion ↓
ResNet152	0.267	<b>0.426</b>
+ mixup	0.224	0.452
+ cutout	0.261	0.451
+ cutmix	<b>0.271</b>	0.474

## 5. おわりに

今回は、ABN にデータ拡張を用いた際の精度比較を行った。ネットワークサイズの変化により、最も性能の良いデータ拡張手法が異なるということが分かった。今後は、これらのデータ拡張手法を組み合わせた際の影響について調査する。

### 参考文献

- [1] H. Fukui, *et al.*, "Attention Branch Network: Learning of Attention Mechanism for Visual Explanation", CVPR, 2019.