

画像中の自己相似性を用いた物体認識・物体検出手法の検討

鹿島大河[†]鶴見智[‡]群馬工業高等専門学校 専攻科生産システム工学専攻[†] 電子情報工学科[‡]

1 初めに

近年、深層学習は画像処理の分野で大きく成功したが [1]、規模が大規模になるほど必要となるデータサイズは大きくなってしまふ。そこで本研究では画像に対して高い圧縮率を持つ、フラクタル符号化に着目した。フラクタル符号化は自己相似性を利用した圧縮手法であり、符号化の際に得られる分割情報自体が画像を特徴づけることができる。この分割情報を利用した深層学習による物体認識および物体検出手法を提案する。

2 原理

2.1 フラクタル符号化

フラクタル符号化とは画像中の自己相似性を利用した符号化手法であり、画像中のある領域 R_i (レンジブロック) を、その 2 倍の大きさである別の領域 D_i (ドメインブロック) に対してアフィン変換を施したものと置き換えるといった処理を繰り返す画像圧縮技法である [2]。 R_i と D_i の相似性 (類似度) を計算し、この類似度が閾値以下になったら R_i を D_i として近似するといった処理を繰り返すことで圧縮される。このフラクタル符号化は符号化の際のパラメータ自体が画像の自己相似性を表し、極めて高速な復号が可能であるといった特徴もある。本研究では画像サイズを 256x256 にリサイズした上でフラクタル符号化を行い、画像に対してどのようにブロックを分割して近似が行われたのかという「分割情報」を利用していくこととする。

2.2 物体認識へのフラクタル分割情報の適用

フラクタル符号化の分割情報を数値的に扱うためにブロックのサイズに応じた bit を考える。それぞれのブロックのサイズ (分割の深さ) と bit の桁を対応させて、対象の bit を 1 とすることで相似性を保持した状態で数値を割り当てることが可能となる。図 1 に画像データそのものと分割ブロックに対して実際に数値を割り当てたものを示す。



図 1: 左: 画像データ 右: 分割ブロック

この時、数値とブロックサイズは対応しているため、自己相似性を失うことなく物体認識へと適用ができる。また、画像の色情報ではなく分割情報のみを扱うために通常よりも小さなデータサイズとして深層学習への入力が可能となる。

2.3 物体検出へのフラクタル分割情報の適用

2.3.1 学習

本研究では fast rcnn[3] をベースとして、物体領域候補にフラクタル分割情報を利用することとする。分割情報はエッジ部分ではサイズが小さくなるという特徴があり、実際に物体検出の候補として意味ある領域はかなり少なくなる。そこで、2x2 のブロックには 32x32 を、8x8 のブロックでは 64x64・・・といった分割情報のブロックに比例したサイズを割り当て、このブロックから一定以上の大きさをもつブロックのみを物体領域候補として採用することで大きく計算量を減らすことができる。また分割ブロックは必ず正方形となるが実際の物体は様々な形状があり得るため、物体領域候補のスケールおよびアスペクト比を変更したものに

Study on Object Recognition and Detection Using Self-Similarity in Image

[†] Taiga Kashima, Advanced Production System Engineering Course, National Institute of Technology, Gunma College

[‡] Satoshi Tsurumi, Department of Information and Computer Engineering

対して roi pooling を行って結合し、畳み込み層を重ねることで位置の微調整を行う。図 2 にこの処理を示す。

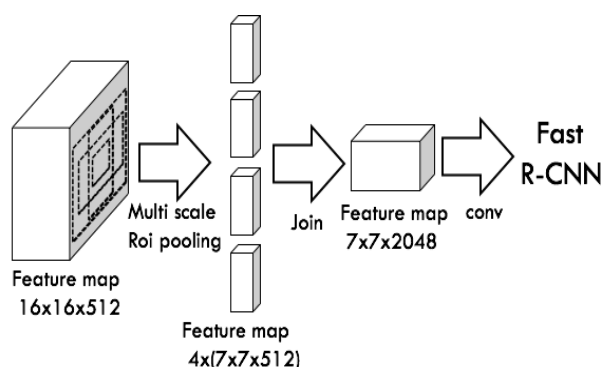


図 2: roi pooling の結合

この方法では様々な形状の roi を結合しているために物体領域とその周辺を取り出すことができ、物体周辺のコンテキストを考慮して物体領域の推定を行うことができる。

2.3.2 推論

小さな物体にも対応するために学習時とは異なる手順で高速に物体領域候補を求める。まず画像をあらかじめ正方形のブロックで分割を行う (fractal block)。次に fractal block の上にフラクタル分割情報を投射し、IoU が一定値以上になった fractal block に関しては結合を行い、新たな fractal block とする。この手順によって生成されたブロックは画像の自己相似性に基づき自動でサイズが定まり、一枚の画像中であっても場所によって異なるサイズのブロックを得ることができる。本実験では様々な大きさの物体に対応するために、初めの fractal block のサイズを 32x32 および 64x64 としたものを全て物体領域候補とした。

3 実験

3.1 物体認識

物体認識では caltech101 を利用した。深層学習を用いて物体認識を行うが、学習には多くのデータを必要ため 100 枚以上の画像が存在する合計 10 種類、3044 枚を用いた。この 3044 枚中 2439 枚を学習データ、605 枚をテストデータとして実験を行った。元画像の大きさは 256x256x3 のデータのため、得られる分割情報の大きさは 128x128x1 の画像と

して考えることができる。この 128x128x1 の分割情報を CNN を用いて GPU 上で学習させた結果、その精度は 96% であった。

3.2 物体検出

分割情報の中でブロックサイズが 8x8 以上の領域を fractal block に投射して物体領域を求めて、vgg16 に対して fine tuning を施し実験を行った。学習データには pascal voc の 2012 を利用し、テストデータには 2007 を使用した。図 3 に実際の検出結果と分割領域を元とした物体候補の例を示す。この実験の結果、テストデータに対して 50mAP を達成し、領域抽出後の入力から non-maximum-supression(nms) まで含めて平均 5FPS を達成した。実装の一部が CPU 上であることや、計算の冗長性があるなどの問題点を改善することによってさらなる高速化が期待できる。

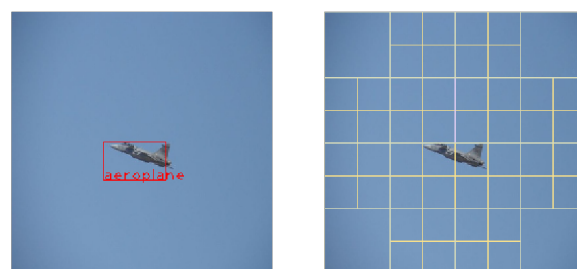


図 3: 左: 検出結果 右: 分割情報

4 まとめ

本研究では画像に対してフラクタル符号化を適用し、得られた分割情報を利用した物体認識手法および高速な物体検出手法を提案した。現在のフラクタル符号化プログラムでは許容近似誤差などのパラメータをヒューリスティクスに求めているので、符号化をより詳細に行うことや任意の画像サイズでの符号化の適用が今後の課題である。また、物体検出における mAP の向上も今後の課題である。

参考文献

- [1] A. Krizhevsky, I. Sutskever, and G. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. In *NIPS*, 2012.
- [2] Y. Fisher. *Fractal Image Compression*. In *SIGGRAPH*, 1992.
- [3] R. Girshick. *Fast R-CNN*. In *ICCV*, 2015.