

## 逆畳み込みニューラルネットワークを用いた輪郭検出

## Contour Detection with Deconvolutional Networks

味曾野 雅史 \*1

Masanori Misono

中山 浩太郎 \*2

Kotaro Nakayama

松尾 豊 \*2

Yutaka Matsuo

\*1 東京大学工学部

Faculty of Engineering, The University of Tokyo

\*2 東京大学大学院工学系研究科

School of Engineering, The University of Tokyo

In this paper, we propose new neural networks for contour detection. The networks use deconvolutional networks to enlarge spatial features instead of linear interpolations for better reconstruction of the location information. Proposed networks achieved state-of-the-art results in the BSDS500 and DRIVE dataset.

## 1. はじめに

輪郭検出問題は、自然画像の中から物体の領域を分ける線を抽出する問題であり、長年研究が続けられている。2015年現在最も輪郭検出の性能が高いとされる HED[1] は Fully Convolutional Network(FCN) [2] と呼ばれる全結合層が存在しない畳み込みニューラルネットワーク (CNN) を利用した輪郭検出器である。HED は画像を入力として受け取りその輪郭線を出力する構造になっており、既存の各ピクセルごとに処理する方法に比べ高速な上、精度の面でも既存手法を大きく上回っている。輪郭検出と近い分野の問題であるセマンティックセグメンテーション (ピクセルごとのクラス分類) でも CNN を応用した手法が近年最も良い精度を出している。その中で逆畳み込みおよびアンプリーングから構成される逆畳み込みニューラルネットワーク [3] が物体境界の正確な位置特定に役立つことが複数の研究で示されている [4, 5]。一方で逆畳み込みニューラルネットワークを輪郭検出に応用した研究例は筆者の知る限りまだ存在しない。本研究では逆畳み込みニューラルネットワークを利用した新たな輪郭検出器を提案し、輪郭検出における逆畳み込みニューラルネットワークの評価をおこなった。

## 2. 問題の定式化及び提案手法

## 2.1 問題の定式化

輪郭線検出問題を教師有り学習の問題として定義する。訓練画像集合を  $S = \{(X_n, Y_n), n = 1, \dots, N\}$  とする。ここで、 $X_n = \{x_j^{(n)}, j = 1, \dots, |X_n|\}$  は輪郭抽出の対象となる 1 枚の画像、 $Y_n = \{y_j^{(n)}, j = 1, \dots, |X_n|, y_j \in \{0, 1\}\}$  は  $X_n$  の輪郭画像であり、 $|X_n|$  は画像のピクセル数 (縦 × 横)、 $y_j$  が 1 のときそのピクセルが輪郭線であることを示す。ある関数  $f(X_n; \theta)$  の出力が  $Y_n$  になるよう訓練データを用いてパラメータを調整する。本研究ではこの関数  $f(X_n; \theta)$  として逆畳み込みニューラルネットワークを用いた輪郭検出ネットワークを利用する。

## 2.2 既存手法の問題点

FCN はプーリングで小さくなった空間的な特徴量を入力サイズと同じにするために線形補完による拡大をおこなう。したがってプーリングを 5 回行った場合は 16 倍もの拡大をおこなうこととなり、空間位置の正確な復元は難しい。HED はネットワークの出力を分岐させ、複数の拡大の結果を重ね合わせるこ

とで輪郭検出の精度を向上させているが、重ね合わせの結果ネットワークの出力がぼやける傾向にある。

## 2.3 提案手法

既存手法の問題点を改善し、輪郭検出の精度を向上するため、特徴量の拡大部分に逆畳み込みニューラルネットワークを利用することを考える。図 1 に今回提案するネットワークの構成図を示す。入力には輪郭線を抽出したい画像である。ネットワークは前半の畳み込み部分および後半の逆畳み込み部分で構成される。畳み込み部分は通常の CNN 同様に畳み込み層およびプーリング層から構成される。後半の逆畳み込み部分では、逆畳み込みあるいは畳み込みおよびアンプリーングを利用して中間の特徴量マップを入力と同じサイズに復元する \*1。逆畳み込みは畳み込みの順伝播と逆伝播を反対にした処理である (図 2)。また、アンプリーングは畳み込み部分でおこなわれたアンプリーングのインデックス情報をもとに、プーリングで反応した部分に特徴量を拡大させる処理である (図 3)。最後の出力層ではシグモイド関数を利用して、各ピクセルごとにそのピクセルが輪郭線であるかどうかの確率を出力する。

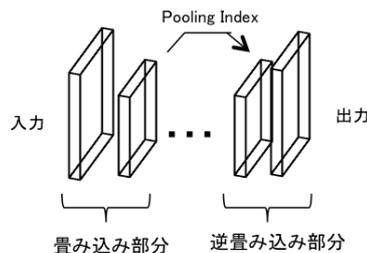


図 1: 提案するネットワークの構成図

最適化対象となる目的関数としては以下の重み付きクロスエントロピー  $\mathcal{L}$  を利用する。

$$\mathcal{L} = -\frac{|Y_-|}{|Y|} \sum_{j \in Y_+} \log P(y_j = 1|X) -$$

\*1 [3] では逆畳み込み層及びアンプリーング層を逆畳み込みネットワークの構成要素として説明している。本論文では便宜上畳み込み層とアンプリーング層を組み合わせた場合も逆畳み込みニューラルネットワークと呼ぶことにする。

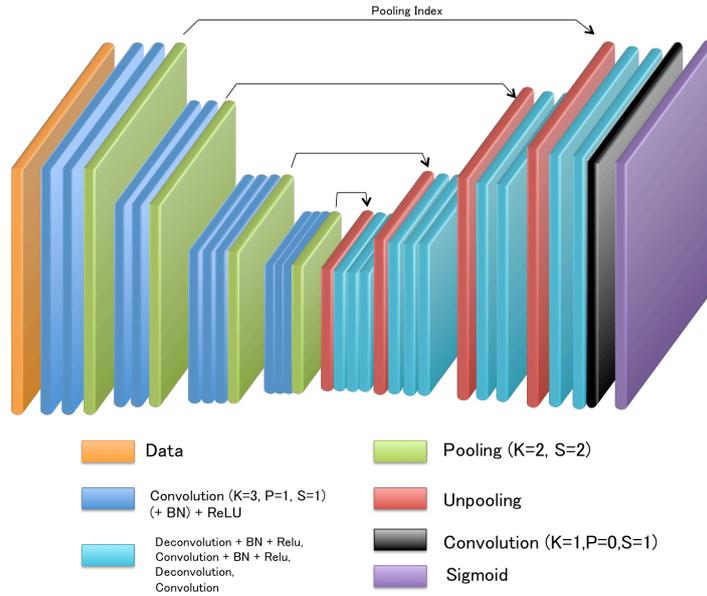


図 4: 評価に利用したネットワーク

$$\left(1 - \frac{|Y_-|}{|Y|}\right) \sum_{j \in Y_-} \log P(y_j = 0|X) \quad (1)$$

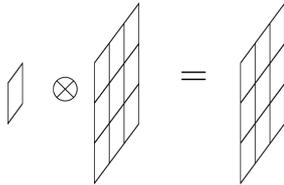


図 2: 逆畳み込み処理

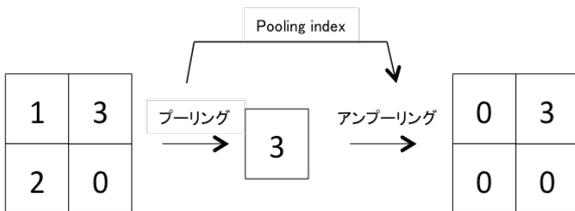


図 3: アンプーリング処理

ここで  $Y_+$  は輪郭線であるピクセルの集合,  $Y_-$  はその逆であり,  $\log P(y_j = 1|X)$  はシグモイド関数の出力である。

提案手法では逆畳み込みニューラルネットワークを利用して拡大させることで, [1] よりも正確な輪郭検出をネットワークの出力として得ることを試みる。またこのネットワークは全結合層が存在しないため, 可変長の入力に対応しており, ネットワーク入先に先立って画像をリサイズしたり切り出す必要はない。

### 3. 実験

#### 3.1 実装

実装は [1] 及び [4] が公開している実装 \*2\*3 を改良する形でおこなった。いずれも代表的な Deep Learning のフレームワークの一つである Caffe[6] をベースにしている。

具体的なネットワーク構成は, VGGNet[7] の 4 回目のプーリングまでのネットワークを取り出し, 後半はそれを反転させて繋げた形にした。実験では後半部分で逆畳み込みを利用する場合 (*deconv*) および畳み込みを利用する場合 (*conv*), さらに後半部分で ReLU 活性化関数を使用しない場合 (*deconv\_norelu*, *conv\_norelu*) の 4 種類のネットワークを評価した。目的関数の最適化には確率的勾配降下法 (SGD) を用い, ILSVRC で訓練済みモデルをファインチューニングした。deconv, conv では各畳み込みおよび逆畳み込み層の後にバッチ正規化 [8] をおこなっている。図 4 に今回実験で利用したネットワークの構成図を示す。

#### 3.2 BSDS500 データセット

BSDS500 データセット [9] は, 輪郭検出およびセグメンテーションのアルゴリズム性能の評価に広く用いられている自然画像のデータセットである。データセットには 300 枚の訓練画像, 100 枚の評価画像, 100 枚のテスト画像が含まれる。学習には [1] が公開している, 回転・拡大により 32 倍のデータ拡張を施したデータを利用した。また, BSDS500 データセットは 5 人によるアノテーション画像が正解画像であるが, [1] と同様に 3 人以上がアノテーションした線を正解画像として利用した。ハイパーパラメータは評価画像を用いてランダムサーチによりモデルごとに調節した。モデル *deconv* のハイパーパラメータ値は "base\_lr" が  $1e-5$ , "lr\_policy" が "step", "iter\_size" が 10, "step\_size" が 5000, "gamma" が 0.1, "weight\_decay" が 0.0002, "momentum" が 0.90, 訓練回数は 6000 Iteration である。

\*2 <https://github.com/s9xie/hed>

\*3 <https://github.com/HyeonwooNoh/caffe>

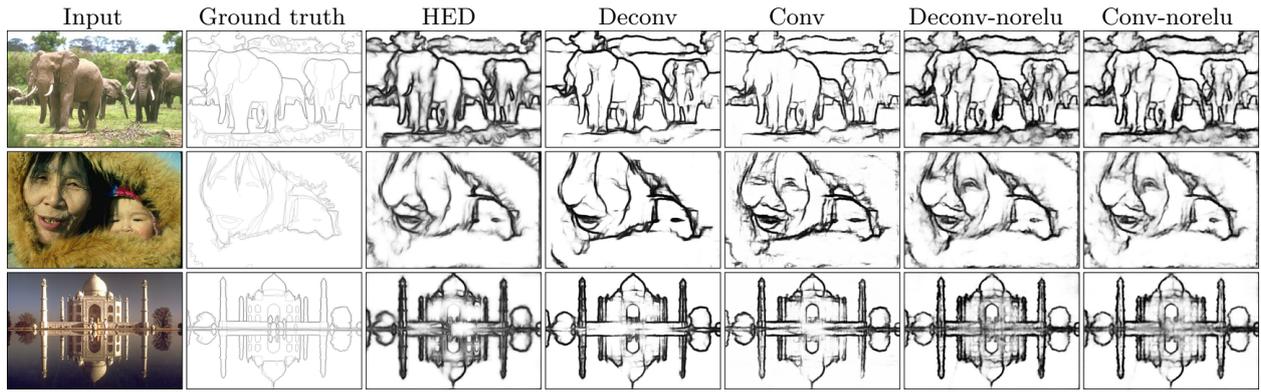


図 6: BSDS500 評価例

表 1 に、提案手法を評価した結果を示す。また図 5 に提案手法の適合率-再現率曲線を、図 6 に実際の評価例の画像を示す。

表 1: BSDS500 テスト結果比較  
(提案手法以外の結果は [1] より引用)

手法	ODS	OIS	AP
Human	.80	.80	-
DeepNets[10]	.738	.759	.758
N <sup>4</sup> -Fields[11]	.753	.769	.784
DeepEdge[12]	.753	.772	.807
CSCNN[13]	.756	.775	.798
DeepContour[14]	.756	.773	.797
HED (Side-output 4)[1]	.740	.759	.672
HED (merged result)[1]	.782	.804	.833
HED (merged result) (arxiv)[15]*4	.788	.808	.840
提案手法 (deconv)	.773	.794	.814
提案手法 (conv)	.767	.787	.810
提案手法 (deconv-norelu)	.783	.801	.753
提案手法 (conv-norelu)	.784	.803	.748
提案手法 (merged)	.795	.814	.827

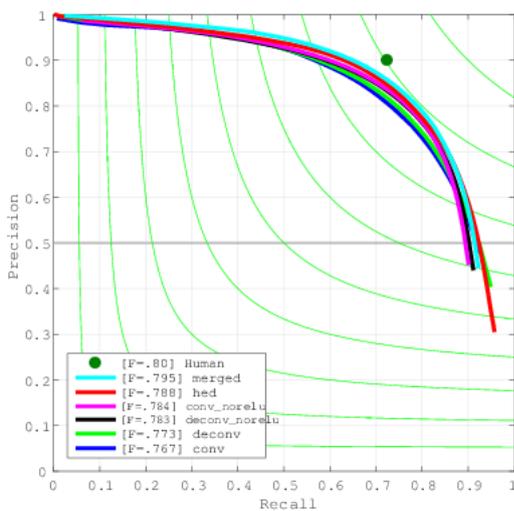


図 5: BSDS500 データセット 適合率-再現率曲線 (F 値は ODS)

ODS はデータセット全体で一つ閾値を決めた時の F 値の最大値, OIS は画像ごとに最適な閾値を決めた場合の F 値の平均値, AP は全体の平均適合率であり, 特に ODS が BSDS500 データセットでの性能指標として用いられている. 評価に先立ち [1] と同様に提案手法で得られた出力に [16] が公開しているツール\*5 を利用し Non-maximum suppression(NMS) 及び細線化処理を施している.

”HED (Side-output 4)”は [1] において複数の重ね合わせをおこなわず, 4 回のプーリングの後に線形補完で拡大して得られた結果である. それと比較するといずれも提案手法として評価したモデルは性能が向上しており, HED と比べても近い精度が出ている. また”提案手法 (merged)”は今回評価した 4 つのモデルの出力を重ね合わせた結果であり, 複数のモデルを重ね合わせることで精度が向上することが確認された.

### 3.3 DRIVE データセット

DRIVE[17] は網膜画像からの輪郭検出の研究をおこなうためのデータセットであり, 20 枚の訓練画像および 20 枚のテスト画像が含まれる. タスクとしてはピクセルごとにそのピクセルが血管であるかどうかを判定することになるので, 輪郭線検出のアルゴリズムがそのまま適用できる. 画像から血管を正確に抽出するには, BSDS500 データセット以上に細かな線の抽出が求められる. 表 2 に DRIVE データセットでの評価結果を示す. ハイパーパラメータは BSDS500 データセットと同じものを用い, 評価は AUC でおこなった.

表 2 に評価結果を, 図 7 に適合率-再現率曲線を示す. 実験結果より, 提案手法は既存の手法に匹敵する精度を出した. BSDS00 と比べると, *deconv-norelu* や *conv-norelu* よりも *deconv* や *conv* の方が AUC が大きくなった.

## 4. 考察

BSDS500 データセットおよび DRIVE データセットにおける評価実験で提案手法は最新手法に匹敵する精度を示しており, 逆畳み込みニューラルネットワークは輪郭検出に有効であるといえる. 細かな線の抽出が特に重要となる DRIVE データセットでは提案手法の *conv*, *deconv* が HED よりも良い精度を出した. これは単純な線形補完よりも逆畳み込みニューラルネットワークを利用した方が細かな線が抽出できたからだと考えられる.

\*4 データ拡張を施し再実験した結果を提出している

\*5 <https://github.com/pdollar/edges>

\*4 Xie らは HED の論文を ICCV'15 に提出した後, arxiv.org に

表 2: DRIVE データセット テスト結果比較  
(提案手法および HED 以外の結果は [11] より引用)

手法	AUC
Becker et al.[18]	.89
N <sup>4</sup> -Fields [11]	.89
HED (fusion-output)[1]	.870
提案手法 (deconv)	.888
提案手法 (conv)	.890
提案手法 (deconv-norelu)	.831
提案手法 (conv-norelu)	.850

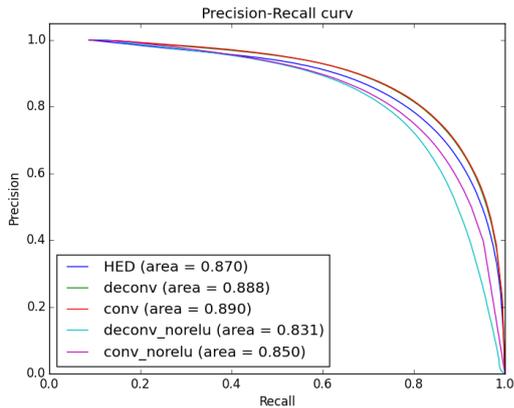


図 7: DRIVE データセット 適合率-再現率曲線

今回実験ではネットワーク後半部分で畳み込み層を使う場合と逆畳み込み層を使う場合の 2 種類を試したが、その二つで大きな差は無かった。一方で ReLU を入れる場合と入れない場合で、入れない方が BSDS500 の ODS 値は良くなったものの、ネットワークの出力としては全体としてぼやけたような画像が生成された (図 6)。DRIVE データセットでは ReLU を入れない方が精度がよかったが、これは DRIVE データセットでは NMS や細線化をおこなっていないので、ネットワークの出力はぼやけない方がよりよいからだと考えられる。

## 5. まとめ

本研究では逆畳み込みニューラルネットワークを用いた新たな輪郭検出手法を提案し、その評価をおこなった。実験結果より、逆畳み込みニューラルネットワークの輪郭検出における有用性が示された。提案手法は HED と比較すると線形補完の代わりに逆畳み込みニューラルネットワークを利用することで、ぼやけの少ない輪郭線の抽出を試みる。

輪郭線検出の精度を上げるためには、ネットワークの出力として得られた画像に条件付き確率場など他の画像処理手法を組み合わせることも重要な研究課題の一つである。一方で筆者はネットワークを改善することでネットワークの出力としてさらに良い結果を得ることができるのではないかと考えている。本研究は輪郭検出ネットワークの構造に関して一つの示唆を与えるものである。

## 参考文献

- [1] S. Xie et al. Holistically-nested edge detection. In *ICCV*, 2015.
- [2] J. Long et al. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
- [3] M.D. Zeiler et al. Deconvolutional networks. In *CVPR*, 2010.
- [4] H. Noh et al. Learning deconvolution network for semantic segmentation. In *ICCV*, 2015.
- [5] V. Badrinarayanan et al. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.
- [6] Y. Jia et al. Caffe: Convolutional architecture for fast feature embedding. In *ACMMM*, 2014.
- [7] K. Simonyan et al. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [8] S. Ioffe et al. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.
- [9] P. Arbelaez et al. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 2011.
- [10] J. Kivinen et al. Visual boundary prediction: A deep neural prediction network and quality dissection. In *AISTATS*, 2014.
- [11] Y. Ganin et al. N<sup>4</sup>-fields: Neural network nearest neighbor fields for image transforms. In *ACCV*, 2014.
- [12] G. Bertasius et al. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In *CVPR*, 2015.
- [13] J. Hwang et al. Pixel-wise deep learning for contour detection. *arXiv preprint arXiv:1504.01989*, 2015.
- [14] W. Shen et al. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *CVPR*, 2015.
- [15] S. Xie et al. Holistically-nested edge detection. *arXiv preprint arXiv:1504.06375*, 2015.
- [16] P. Dollár et al. Fast edge detection using structured forests. *IEEE TPAMI*, 2015.
- [17] J. Staal et al. Ridge-based vessel segmentation in color images of the retina. *IEEE T-MI*, 2004.
- [18] C.J. Becker et al. Supervised feature learning for curvilinear structure segmentation. In *MICCAI (1)*, 2013.