

Deep Sign Retrieval: 位相回復と深層学習に基づく DCT 係数に伴う符号情報の圧縮 Compression of Sign Information of DCT Coefficients via Deep Sign Retrieval

鈴木 圭* 都竹 千尋* 高橋 桂太* 藤井 俊彰*
Kei Suzuki Chihiro Tsutake Keita Takahashi Toshiaki Fujii

1 はじめに

JPEG などの離散コサイン変換 (DCT) を利用する画像圧縮において、DCT 係数に伴う符号ビットの発生確率が等しく、その圧縮が課題となっている。この問題を解決するために、著者らは Phase Retrieval (PR) の理論に基づいて、DCT 係数の絶対値からその符号を復元する Sign Retrieval (SR) 手法を提案している [1]。SR では、画像がウェーブレット領域でスパース表現できると仮定し、ウェーブレット縮退を適用する。この方法では、エッジ成分やフラットなブロックの符号情報を高精度に復元できる一方、テクスチャのようなアクティビティの高いブロックにおいてはスパースモデルが破綻するため、復元した符号情報に誤差が生じる。そこで、本研究では Algorithm Unrolling [2] に基づいて、ウェーブレット縮退処理を Deep Neural Network (DNN) の推論に置き換えた Deep Sign Retrieval (DSR) 手法を提案する。画像の解析的なモデルとして用いられてきたウェーブレットとは異なり、DNN はテクスチャのような画像固有のパターンを学習モデルによって効率的に表現できるため、画像モデリングの意味でウェーブレットを凌駕する。JPEG および SR と比較して、レート歪性能の観点から DSR の有効性を検証する。

記法: 以降では、二次元画像を $\mathbf{x} \in \mathbb{R}^{N_1 \times N_2}$ 、その画素値を x_{m_1, m_2} と表記する。 \mathbf{x} に含まれる 8×8 画素ブロックを $\mathbf{x}_{b_1, b_2} \in \mathbb{R}^{8 \times 8}$ 、その画素値を $x_{b_1, b_2; i_1, i_2}$ と表記する。ここで、 b_1, b_2 はブロックのインデックス、 i_1, i_2 はブロック内のインデックスを表す。 \mathbf{x} と \mathbf{x}_{b_1, b_2} の DFT および DCT を各々 $\mathbf{x}^{\text{dft}} \in \mathbb{C}^{N_1 \times N_2}$ 、 $\mathbf{x}_{b_1, b_2}^{\text{dct}} \in \mathbb{R}^{8 \times 8}$ 、これらの変換係数を $x_{m_1, m_2}^{\text{dft}}$ 、 $x_{b_1, b_2; u_1, u_2}^{\text{dct}}$ と表し、 m_1, m_2 と u_1, u_2 は各々 DFT および DCT 領域のインデックスである。

2 関連研究

2.1 Phase Retrieval と Phase Max 法

PR は古典的な信号復元問題であり、非凸最適化問題

$$\text{Find } \text{phase}(x_{m_1, m_2}^{\text{dft}}) \text{ from } |x_{m_1, m_2}^{\text{dft}}|, \forall m_1, m_2 \quad (1)$$

によって定式化される。式 (1) は NP-hard 問題であることが知られており、有限時間内での求解が不可能である。そこで、文献 [3] では式 (1) を凸緩和した最適化問題

$$\max_{\mathbf{z} \in \mathbb{R}^{N_1 \times N_2}} \langle \phi, \mathbf{z} \rangle \text{ s.t. } |z_{m_1, m_2}^{\text{dft}}| \leq |x_{m_1, m_2}^{\text{dft}}|, \forall m_1, m_2 \quad (2)$$

を解いている。 $\phi \in \mathbb{R}^{N_1 \times N_2}$ はアンカーベクトルと呼ばれ、原画像 \mathbf{x} と ϕ のコサイン類似度が十分に大きく、かつ $|x_{m_1, m_2}^{\text{dft}}|$ がオーバーサンプリングされている場合*1、式 (1) と式 (2) の解は完全に一致する。なお、式 (2) は Phase Max 法と呼ばれ、線形計画法によって多項式時間で求解できる。

2.2 文献 [1] の手法

2.2.1 符号化器と復号器

著者らは DCT 係数の符号情報を効率的に圧縮するために、下記の符号化器および復号器を構築しており、符号化器と復号器のベースラインとして JPEG を採用している。以降では、JPEG 符号化画像を \mathbf{y}

と表す。文献 [1] では $\text{sgn}(y_{b_1, b_2; u_1, u_2}^{\text{dct}})$ の効率的なロスレス復号を目的としており、最終的に得られる画像は \mathbf{y} であることに注意されたい。

符号化器: 全ての交流成分の符号情報を JPEG のビットストリームから取り除く、すなわち絶対値 $|y_{b_1, b_2; u_1, u_2}^{\text{dct}}|$ を符号化する。次に、次節で述べる SR 問題を解くことによって、 $|y_{b_1, b_2; u_1, u_2}^{\text{dct}}|$ の符号をローカル予測する。最後に、次式に示す誤り訂正ビット列を符号化する。

$$e_{b_1, b_2; u_1, u_2} = \text{sgn}(y_{b_1, b_2; u_1, u_2}^{\text{dct}}) \oplus \text{ret_sgn}_{b_1, b_2; u_1, u_2} \quad (3)$$

ここで、 $\text{ret_sgn}_{b_1, b_2; u_1, u_2}$ は SR によって復元された符号、 \oplus は XOR オペレータであり、 $\text{ret_sgn}_{b_1, b_2; u_1, u_2}$ が $\text{sgn}(y_{b_1, b_2; u_1, u_2}^{\text{dct}})$ に一致する場合、誤り訂正ビット列のシャノンエントロピーは 0 となり、効率的な圧縮が可能となる。

復号器: まず、絶対値 $|y_{b_1, b_2; u_1, u_2}^{\text{dct}}|$ を復号し、SR 問題を解くことによって $\text{ret_sgn}_{b_1, b_2; u_1, u_2}$ を得る。次に、誤り訂正ビット列 $e_{b_1, b_2; u_1, u_2}$ を復号し、 $\text{ret_sgn}_{b_1, b_2; u_1, u_2}$ との XOR を計算することによって符号情報をロスレス復号する。最後に、ロスレス復号された符号と絶対値 $|y_{b_1, b_2; u_1, u_2}^{\text{dct}}|$ から \mathbf{y} を復号する。

2.2.2 Sign Retrieval

SR とは、式 (1) の PR における位相関数 $\text{phase}(x_{m_1, m_2}^{\text{dft}})$ と振幅関数 $|x_{m_1, m_2}^{\text{dft}}|$ の各々を、符号関数 $\text{sgn}(y_{b_1, b_2; u_1, u_2}^{\text{dct}})$ と絶対値関数 $|y_{b_1, b_2; u_1, u_2}^{\text{dct}}|$ に置き換えた非凸最適化問題

$$\text{Find } \text{sgn}(y_{b_1, b_2; u_1, u_2}^{\text{dct}}) \text{ from } |y_{b_1, b_2; u_1, u_2}^{\text{dct}}|, \forall b_1, b_2, u_1, u_2 \quad (4)$$

のことである。Phase Max 法によると、式 (4) の解はその凸緩和問題を解くことによって得られるが、画像の符号化において $|y_{b_1, b_2; u_1, u_2}^{\text{dct}}|$ がオーバーサンプリング条件を満たすにはオーバーコンプリート DCT の絶対値を送る必要があり、符号量のオーバーヘッドが極めて大きくなる。そこで、通常の DCT 係数をオーバーコンプリート DCT 係数の圧縮サンプリングとみなして、式 (4) の凸緩和問題を正則化した Regularized Sign Max 問題

$$\max_{\mathbf{z} \in \mathbb{R}^{N_1 \times N_2}} \sum_{b_1, b_2} \langle \mathbf{z}_{b_1, b_2}, \phi_{b_1, b_2} \rangle - \lambda \|\psi(\mathbf{z})\|_1 \text{ s.t. } |z_{b_1, b_2; u_1, u_2}^{\text{dct}}| \leq |y_{b_1, b_2; u_1, u_2}^{\text{dct}}|, \forall b_1, b_2, u_1, u_2 \quad (5)$$

を解く。ここで、 λ は正則化パラメータ、 ψ はタイトフレームウェーブレット変換であり、式 (5) より得られる解の DCT 係数の符号が $\text{ret_sgn}_{b_1, b_2; u_1, u_2}$ である。

式 (5) は Fienup の反復解法 [4] を用いて、

$$\mathbf{f}_{[\theta+1]} = \psi^{-1} \left(\text{sign}(\psi(\mathbf{z}_{[\theta]})) (|\psi(\mathbf{z}_{[\theta]})| - \lambda)_+ \right) \quad (6)$$

$$\mathbf{g}_{[\theta+1]} = \mathbf{f}_{[\theta+1]} + \frac{1}{\mu} \phi \quad (7)$$

$$\mathbf{z}_{[\theta+1]} = \text{proj}_{\text{dct}} \left(\mathbf{g}_{[\theta+1]} \right), \quad (8)$$

によって解ける。 θ は反復回数、 $(a)_+ = \max(a, 0)$ であり、式 (6) はウェーブレット縮退、 proj_{dct} は θ 回目の更新における \mathbf{g} の DCT 係数 $g_{b_1, b_2; u_1, u_2}^{\text{dct}}$ を凸集合 $[-y_{b_1, b_2; u_1, u_2}^{\text{dct}}, y_{b_1, b_2; u_1, u_2}^{\text{dct}}]$ に射影する凸射影作用素である。

* 名古屋大学 大学院工学研究科 情報・通信工学専攻

*1 m_1, m_2 が小数精度でサンプリングされていること。

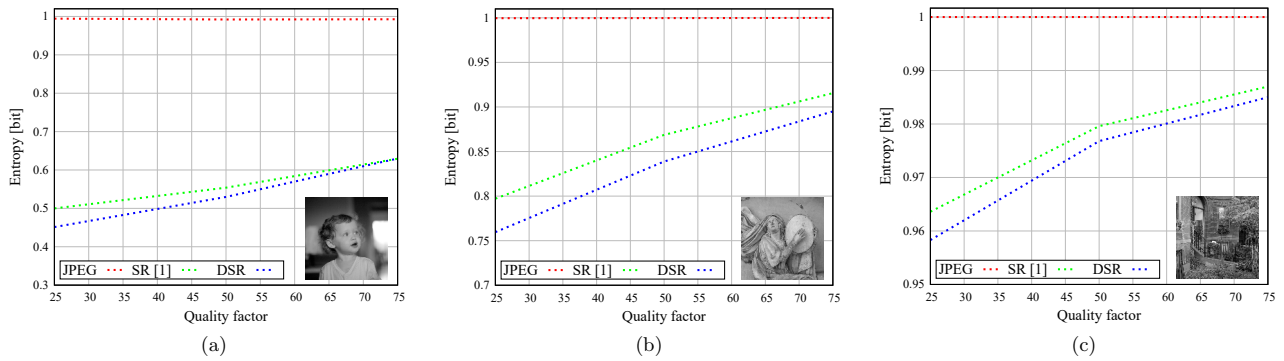


図1: 各手法における符号あたりのエントロピー

表1: ネットワーク構造

Layer	Kernel size	In ch	Out ch	Input
Conv1	3	1	32	input image
Conv2	3	32	32	Conv1
Conv3	3	32	32	Conv2
⋮	⋮	⋮	⋮	⋮
Conv7	3	32	32	Conv6
Conv8	3	32	1	Conv7

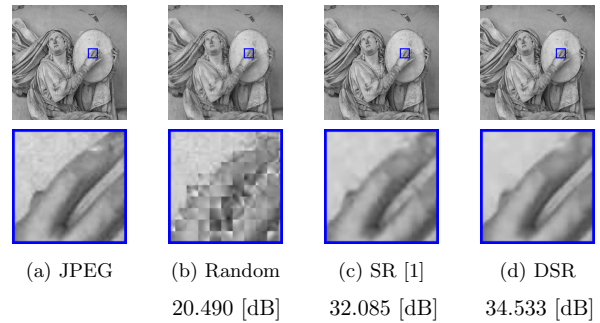


図2: 復元画像の例

3 Deep Sign Retrieval

ウェーブレット縮退がアフィン変換および非線形関数によって構成されていることに着目すると、これはDNNにおける重み乗算、バイアス加算、および活性化関数と同等の手続きとなっている。Algorithm Unrolling [2]では、Fienupの解法のような反復アルゴリズムに含まれる非線形処理をDNNに置き換えることによって、解の精度を向上するものであり、DSRでは式(6)を後述する方法で学習したDNNに置き換える。

$$\mathbf{f}_{[\theta+1]} = \text{DNN}_{[\theta]}(\mathbf{z}_{[\theta]}) \quad (9)$$

ここで、 $\text{DNN}_{[\theta]}$ は反復回数 θ ごとに個別に構築されるDNNである。

表1は $\text{DNN}_{[\theta]}$ のネットワーク構造を表している。Convは二次元畳み込み層である。学習には、CLIC2020で公開されている1290枚の画像をグレースケール化し、 $N_1 \times N_2 = 512 \times 512$ にリサイズしたデータを用いた。まず、 $\theta = 0$ の学習では、ブロックごとに直流のみをもつ画像を入力とし、 $\text{DNN}_{[0]}$ の出力がJPEG符号化画像と等しくなるように、5エポック学習した。また、 $\theta > 0$ の学習では、前ステップの $\text{DNN}_{[\theta-1]}$ が推論した復元画像(1290枚)のDCT係数に式(8)の凸射影を適用した $\mathbf{z}_{[\theta]}$ を入力とし、出力がJPEG符号化画像と等しくなるように、5エポック学習した。 $\text{DNN}_{[\theta]}$ は θ 毎に異なるネットワークパラメータを有するようにした。なお、全てのDNNの最適化には学習率0.001のAdamを用いた。

4 実験

学習データセットに含まれない $N_1 \times N_2 = 512 \times 512$ のグレースケール画像3枚を原画像として、JPEG圧縮の際の量子化パラメータQFを25, 50, 75と変化させ、JPEG, SR [1], およびDSRの符号あたりのエントロピーを比較した。このエントロピーは、JPEGではDCT係数の符号が持つ情報量に、SRおよびDSRでは誤り訂正ビット列が持つ情報量に対応する。なお、SRとDSRでは、復元処理の後に誤り訂正ビット列を用いて符号を完全に復元するため、最終画像はJPEG符号化画像に一致する。DSRでは、正規化パラメータ $\lambda = 1$ 、反復回数は300として、300個の $\text{DNN}_{[\theta]}$ を個別に学習した。

実験結果を図1に示す。(a)のQF = 75を除いてDSRが最も低い

エントロピーを達成しており、これはJPEG符号化画像が本来持っていた係数符号を直流成分から高精度に推定できていることを示している。また、図1(b),(c)に着目すると、アクティビティの高いブロックを多く含む画像に対してDSRが特に有効であると言える。図2に復元画像の例を示す。(a)はJPEG符号化画像、(b)は符号をランダムに割り当てた画像、(c)と(d)は各々SRおよびDSRによる復元画像(誤り訂正前)であり、JPEG符号化画像に対するPSNRも記載した。この結果から、DSRは極めて高精度にJPEG符号化画像を復元できていることがわかる。これは誤り訂正ビット列のエントロピーの低減につながるため、図1の結果は妥当である。さらに、DSRでは明らかに高周波成分を復元できていることから、画像モデリングの意味でDNNがウェーブレットを凌駕すると言える。

5 まとめ

本稿では、PRとDNNに基づいて、DCT係数の符号情報を圧縮するDSR手法を提案した。今後の展望は、 $\text{DNN}_{[0]}$ に対して超解像ネットワークなどの高度なモデルを組み込むことである。

参考文献

- [1] C. Tsutake, K. Takahashi, and T. Fujii, "An efficient compression method for sign information of DCT coefficients via sign retrieval," *IEEE Int. Conf. Image Process. (ICIP)*, 2021 (accepted).
- [2] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.
- [3] T. Goldstein and C. Studer, "Phasemax: convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2675–2689, 2018.
- [4] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 23, no. 15, pp. 2758–2769, 1982.