Vol.2013-CVIM-187 No.30
2013/5/30

ディープボルツマンマシンを用いた線画の修復

白川 悠太[1,a)]　岡谷 貴之[1,b)]

概要：ディープラーニングの方法は，画像認識の様々なベンチマークテストにおいてよい結果を残しているが，それらは主に畳込みニューラルネットワークを用いた教師あり学習によるものである．一方，画像を学習対象とした無教師学習，特に全結合型の多層ネットワークを用いたものは，必ずしもベンチマークテストで注目されるような結果は残していない．本論文では，ディープボルツマンマシンを用いた多層ネットワークの無教師学習により，線画の生成モデルを学習し線画の欠損修復を行う方法を述べる．その上で自由度を揃え単層のモデルと比較し，全結合型の多層ネットワークの有効性について比較実験を行う．

1. はじめに

物体カテゴリー認識を中心とする画像認識のための方法として，ディープラーニング，すなわちニューラルネットワークを核とする多層のネットワークを用いる学習方法が，最近注目を集めている．現在までに，ILSVRC [1] をはじめとする数々のベンチマークテストで，よい結果 [2] を残している．

現在のディープラーニングのブームは，多層のビリーフネット（DBN）の学習を可能とした方法の発見を端緒とする [3]．DBN は，挙動を確率的に記述するボルツマンマシンの一種で，通常，層間のユニットがすべて結合した多層構造を持つネットワークである．入力データの集合に対し，入力層から順番に各層を無教師で学習するプレトレーニングを行うことで，DBN の無教師学習がうまく行えるようになった．入力サンプル１つに対し，学習後のネットワークにこれを入力したときの上位層の出力をその特徴量として用いて，その識別・分類を行う．音声認識では，この方法が良い成果を挙げている [4]．

しかしながら，画像認識の各ベンチマークで上述のようによい成果を挙げているのは，畳込みニューラルネットワーク（以下 CNN）を用いた，教師あり学習である．CNN は，層間の結合は局所的に限定され，かつ結合重みを上位層のユニットどうしで共有することと，学習の対象とならないプーリング層が埋め込まれていることを特徴とする．教師ありの学習であることと併せて，全結合ネットワークに対する無教師学習とは大きく異なる．

もちろん，画像に対する無教師学習の研究はこれまでに多数行われている．そこでの筋書きは，まず無教師学習によってよい画像特徴を学習し，そうして得た特徴を元に教師あり学習を行なって，認識をうまく実行しようというものである．しかしながらこのやり方では，今のところ CNN には及ばない性能しか実現できていない．実際，無教師学習の研究のほとんどは，どちらかと言えば学術的な関心から行われており，もっぱらどのような画像特徴が学習されたかのみが議論されてきた．例えば，階層的な構造を持つ特徴の学習に成功したという報告 [5] や，「おばあさん細胞」の再現に成功したという報告 [6] が例である．（これらの研究で，ネットワークにはやはりプーリングや畳込みの構造が取り入れられていることは特筆すべきである（例外は視覚野 V2 領域の特徴を無教師学習によって再現したという [7]），）

ディープラーニングとはその名の通り，ネットワークの多層性こそが本質と考える方法といえる．多層の効果は，CNN の場合すでによく実証されているが，（全結合ネットワークの）無教師学習では，それほど明確ではない．（それどころか，多層にするよりも単層にした方が，良い結果を得たという報告 [8] さえある．）

無教師学習では，学習は入力サンプルの集合を忠実に再現するように行われる．学習される特徴に注意を向けがちだが，学習されるのは，入力サンプルそのものを生成できるようなモデル（生成モデル）である．（その意味では DBM や制約付きボルツマンマシン（RBM）などのボルツマンマシンの方法は，これを素直に体現している．）それゆえ，無教師学習の用途は，物体認識—分類に適したよい画像特徴を見つけること—よりもむしろ，サンプルの生成

[1] 情報処理学会
IPSJ, Chiyoda, Tokyo 101–0062, Japan
[†1] 現在，東北大学
Presently with Tohoku Uniersity
[a)] shirakawa@vision.is.tohoku.ac.jp
[b)] okatani@vision.is.tohoku.ac.jp

モデルを有効に使える問題であると考えるのがより自然である．

本研究では，以上を踏まえて，ボルツマンマシンの最も進んだ形とも言えるディープボルツマンマシン（DBM）[9]を用いて，多層のネットワークの無教師学習により，線画の生成モデルを学習し，それを線画の欠損修復に応用することを考えた．特に，画像を対象とした多層のネットワークの無教師学習において，多層であることの利点がかならずしも明確でないことを特に意識し，その利点を厳密に実験評価することを同時に行った．なお [10] では，多層のデノイジングオートエンコーダを用いて自然画像の画像復元を行っているが，そこでは多層性の効果は評価されていない．

線画を対象としたのは主に次の 2 つの理由による．第一に，学習対象の自由度が限定され，それゆえに取り扱いが簡単になることである．自然画像を対象とすると，その自由度の高さから大規模なネットワークが必要となると思われるが，線画の場合はより小規模なネットワークでも学習できるはずで，実験が行い易い．第二に，線画を対象とすることで，むしろ自然画像のより高度な構造を選択的に扱えそうなことである．線画は，自然画像にエッジ検出を施した後の画像に類似する．それゆえ，エッジ検出という低レベルな処理をバイパスし，高レベルな構造を最初から扱っているともいえる．また，線画は，簡単化，抽象化されたものであるにもかかわらず，われわれ人はよく解釈でき，そのこともこれを裏付ける．

2. ボルツマンマシン

本節ではわれわれが扱ったボルツマンマシンについて簡単にまとめつつ，以降の実験に用いた実際の計算方法について述べる．

2.1 ボルツマンマシン

無教師学習を行うにあたり，まず確率分布を表現するためのモデルを仮定する．われわれが仮定するモデル，ボルツマンマシン（Boltzmann Machines）は図 1 の (a) に示されるような相互に対称なエッジで任意の 2 つのノードが関係付けられているネットワーク構造を持つ．そのため，モデル全体の状態をノード自身，およびノード間の関係を考慮しつつ，確率的に記述できる．

ボルツマンマシンのノードは 2 つの集合に分けられ，入力画像の各ピクセルの輝度値に対応する可視ノード v_i と，抽出された画像の特徴量に対応する隠れノード h_j である．各ノードは 0 もしくは 1 の 2 値の状態をとることができる[*1]．つまり，われわれの実験では可視ノードが 1，つまり発火しているというのはそのノードに対応しているピクセルが白であることを示しており，反対に 0 である場合は黒であることを示す．一方で隠れノードが 1 というのはそのノードが学習した特徴が検出されているということを示し，反対に 0 である場合はその特徴が検出されなかったことを示す．また，それぞれの集合で構成される層をそれぞれ可視層 $\mathbf{v}$，隠れ層 $\mathbf{h}$ と呼ぶ．

可視層 $\mathbf{v}$ と隠れ層 $\mathbf{h}$ をもつボルツマンマシンのエネルギーは式 (1) のように記述される．

$$E(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) = -\sum_{ij} v_i W_{ij} h_j - \sum_{ij} v_i S_{ij} v_j - \sum_{ij} h_i T_{ij} h_j - \sum_i b_i v_i - \sum_j c_j h_j \quad (1)$$

ここで $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{S}, \mathbf{T}, \mathbf{b}, \mathbf{c}\}$ であり，エネルギー関数のパラメータを示す．$\mathbf{W}$ は可視ノードと隠れノードの間の結合，$\mathbf{S}$ と $\mathbf{T}$ は可視ノード間，および隠れノード間の結合，$\mathbf{b}$ と $\mathbf{c}$ はそれぞれ可視ノード，および，隠れノードそれぞれ個々に存在するバイアスを示す．

このボルツマンマシンがモデル全体の状態 $\{\mathbf{v},\mathbf{h}\}$ に与える確率分布は式 (2) で表される．ここで式 $Z(\boldsymbol{\theta})$ は分配関数とよばれ，確率分布を正規化するための定数である．

$$p(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\exp(-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})) \quad (2)$$

$$Z(\boldsymbol{\theta}) = \sum_{\mathbf{vh}} \exp(-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) \quad (3)$$

可視ノード，および隠れノードに与えられる条件付き確率は式 (4) および，式 (5) のように与えられる．

$$p(v_i|\mathbf{h},\mathbf{v}_{-i};\boldsymbol{\theta}) = sigm(\sum_j W_{ij} h_j + \sum_{m\setminus i} S_{im} v_m + b_i) \quad (4)$$

$$p(h_j|\mathbf{v},\mathbf{h}_{-j};\boldsymbol{\theta}) = sigm(\sum_i v_i W_{ij} + \sum_{k\setminus j} h_k T_{kj} + c_j) \quad (5)$$

ここで $sigm(x) = 1/(1+\exp(-\mathrm{x}))$ であり．シグモイド関数と呼ばれる非線形な関数である．

これらの式で表されているように，一般的なボルツマンマシンでは注目するノードの値の更新にはそれ以外のノードすべての状態が影響する．そのため，モデルの状態 $\{\mathbf{v},\mathbf{h}\}$ を正確に計算するのが困難である．

2.2 制約付きボルツマンマシン

一般的なボルツマンマシン（Boltzmann Machines）の計算困難性を解消するため，可視ノードと隠れノード間のみ結合をするという制約を加えたモデルを制約つきボルツマンマシン（Restricted Boltzmann Machines : RBM）（図 1 の (b)）と呼ぶ．RBM のエネルギー関数は次式で表される．

$$E(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) = -\sum_{ij} v_i W_{ij} h_j - \sum_i b_i v_i - \sum_j c_j h_j \quad (6)$$

[*1] 簡単な式の変形により連続値に拡張することができるが，問題の簡単化のため以下のわれわれの実験では 2 値のモデルのみをあつかう．

RBM のネットワークは 2 部グラフの構造をとるため，一方の層の状態が決まるともう一方の層の確率分布は式 (7)，および式 (8) により正確に求められる．

$$p(v_i|\mathbf{h};\boldsymbol{\theta}) = sigm(\sum_j W_{ij}h_j + b_i) \tag{7}$$

$$p(h_j|\mathbf{v};\boldsymbol{\theta}) = sigm(\sum_i v_i W_{ij} + c_j) \tag{8}$$

RBM の学習は学習データセットに対する次式の（対数）尤度を最大化するパラメータ $\boldsymbol{\theta}^*$ を求めることで行う．

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_n \log \sum_{\mathbf{h}} p(\mathbf{v}^{(n)}, \mathbf{h}; \boldsymbol{\theta}) \tag{9}$$

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \mathcal{D}(\boldsymbol{\theta}) \tag{10}$$

対数尤度の最大化は，Kullback らにより定義された 2 つの分布間の近さを表す KL 情報量 $\mathcal{D}(\boldsymbol{\theta})$[11] を最小にしていくことと同じである．KL 距離のパラメータについての導関数は式 (11) のように 2 つの項の差に分解できる．

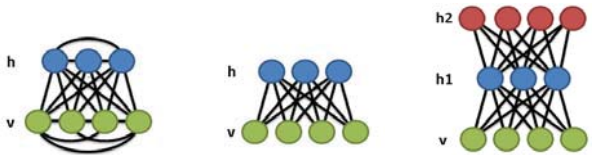
$$\begin{aligned}\frac{\partial \mathcal{D}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{1}{N}\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \sum_{\mathbf{vh}} \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\mathbf{h}|\mathbf{v}) q_0(\mathbf{v}) - \sum_{\mathbf{vh}} \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\mathbf{v},\mathbf{h}) \\ &= \left\langle \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\rangle_{\text{data}} - \left\langle \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\rangle_{\text{model}}\end{aligned} \tag{11}$$

1 つ目の項はエネルギー関数のパラメータについての導関数の経験分布に基づく期待値であり，データ項と呼ぶ．RBM では前述のように片方の層の状態が定まるともう片方の層の確率は正確に計算できる．この性質よりデータ項の計算は正確に行える．一方で 2 つ目の項はエネルギー関数のパラメータについての導関数のモデルの確率分布に基づく期待値であり，モデル項と呼ぶ．この項の計算には分配関数が含まれるため，モデルのとり得るすべての状態についての和の計算が必要となり依然計算が困難である．そのため，モデル項の計算には Contrastive Divergene[12] に代表される近似手法を用いる．以下の実験では Persistent Contrastive Divergence[13] を用いて近似計算を行った．

2.3 スパース正則化

スパースコーディング（Sparse Coding）[14] を筆頭に，画像を少ない基底により表現できる過完備な特徴辞書を学習することは有効であると考えられている．このような認識の下に RBM の学習に対してスパース性を加えたものを疎な制約付きボルツマンマシン（Sparse RBM）[7] と呼ぶ．具体的には，最小化する目的関数に密な発火に対するペナルティを与えるスパース制約項を加えたモデル（式 (12)）である．

$$\begin{aligned}\mathcal{D}(\boldsymbol{\theta}) = &-\frac{1}{N}\sum_n \log \sum_{\mathbf{h}} p(\mathbf{v}^{(n)}, \mathbf{h}; \boldsymbol{\theta}) \\ &+\lambda \sum_j |p - \frac{1}{N}\sum_n \mathbb{E}[h_j^{(n)}|v_j^{(n)}]|^2\end{aligned} \tag{12}$$



(a) (b) (c)

図 1 ボルツマンマシンのグラフィカルモデル．(a) 一般的なボルツマンマシン．(b) 制約つきボルツマンマシン．(c) ディープボルツマンマシン．

スパース制約を加えることで，単にモデル分布と経験分布との距離を縮めるだけではなく，より疎な発火により経験分布との距離を縮めるよう学習する．この制約のもとでモデルの尤度を高めていくと，同じ情報を限られた発火により表現できるようになり，同一パラメータ数においてより表現できる特徴が増すことが期待される．

2.4 ディープボルツマンマシン

より複雑な確率分布を学習するため，RBM を積み重ねたような多層のネットワークを構成するモデルをディープボルツマンマシン（Deep Boltzmann Machines : DBM）（図 1 の (c)）と呼び，そのエネルギー関数は式 (13) のようになる．

$$\begin{aligned}E(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) = &-\sum_{ij} v_i W_{ij}^1 h_j^1 - \sum_{jk} h_j^1 W_{ij}^2 h_k^2 \\ &-\sum_i b_i v_i - \sum_j c_j h_j^1 - \sum_k d_k h_k^2\end{aligned} \tag{13}$$

最下層の可視層に画像の輝度値が与えられ，最上層の隠れ層に画像の特徴が現れる．中間では下方の RBM の隠れ層の状態を上方の RBM の可視層として扱う．学習は，層ごとの逐次的な事前学習をまず行い，その後モデル全体での最適化（微調整）を行う [9], [15], [16].

2.5 学習後のモデルからの画像生成

学習された確率分布 $p(\mathbf{v},\mathbf{h};\boldsymbol{\theta})$ から確率的にサンプリングすることで画像を生成できる．また，生成する際に任意のノードに対して条件を付与することで，その条件のもとの条件付き確率分布を算出することができる．例えば，ある隠れ層のあるノードが 1 で残りのノードが 0 である条件を付与することで，そのモデルのあるノードが反応する特徴を可視化することができる．また，可視層の一部に対して条件を付与することで，残りの部分を推定することもできる．

3. 実験

生成モデルを用いた応用の例として線画の欠損復元を考える．また，多層性の効果を見るため，単層の RBM と 2 層の DBM の性能を，双方の自由度を等しくした上で比較した．

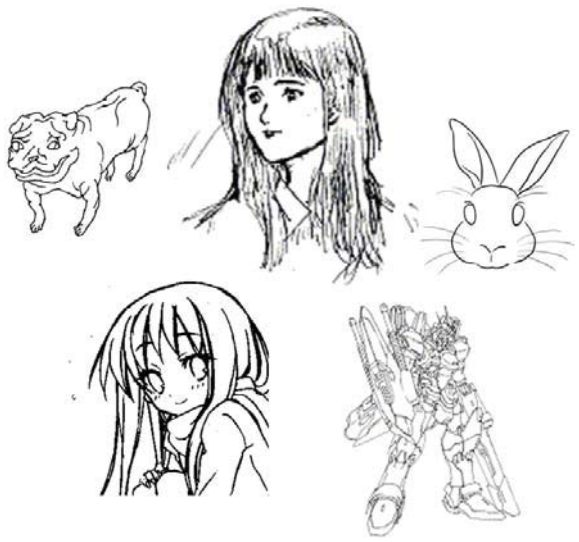


図 2 実験で用いた線画の例.

図 3 線画から切り出した 16×16 画素のパッチの例.

3.1 実験方法

線画は，インターネットより収集した 487 枚の（著作権の放棄された）ものを用いた．図 2 にその例を示す．学習の際には 16 × 16 画素の大きさのパッチを，各線画中のランダムな位置から 5,000 枚切り出し，合計 2,435,000 枚のパッチを生成した．切り取られたパッチの例を図 3 に示す．

以上のデータセットに対し，RBM および 2 層の隠れ層を持つ DBM について，それぞれ 4 つの規模のモデル，合計 8 個のモデルについて学習を行った．各モデルのパラメータ（自由度）の数を表 1 に示す．

さらに，学習されたモデルを用いて欠損を含む画像の復元を行った．なお，復元を欠損の位置は既知であるとし，次の手順で行う．はじめに，欠損を含む画像に対し，学習時に使用したパッチと同サイズの領域を，重なりを許しながら走査し，各領域において欠損を受けていない部分の輝度値が与えられた条件での条件付き確率分布を求め，欠損を受けた部分の輝度値の期待値を計算する．画像全体を走査し終えた後，各領域の期待値の平均を求め，各ピクセルの期待値として採用する．最後に，採用された期待値に従い欠損部分の輝度値（0/1）をランダムにサンプリングし，欠損部分の値とする．

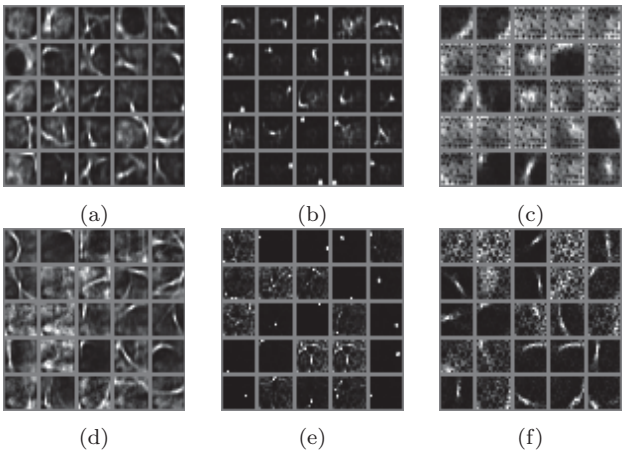


図 4 学習後の RBM および DBM の各層の隠れノードにより抽出される特徴．(a) 隠れノードの数が 100 の RBM．(b) 各隠れ層に 75 の隠れノードをもつ DBM の 1 層目．(c) 各隠れ層に 75 の隠れノードをもつ DBM の 2 層目．(d) 隠れノードを 360 もつ RBM．(e) 各隠れ層に 200 の隠れノードをもつ DBM の 1 層目．(f) 各隠れ層に 200 の隠れノードをもつ DBM の 2 層目．

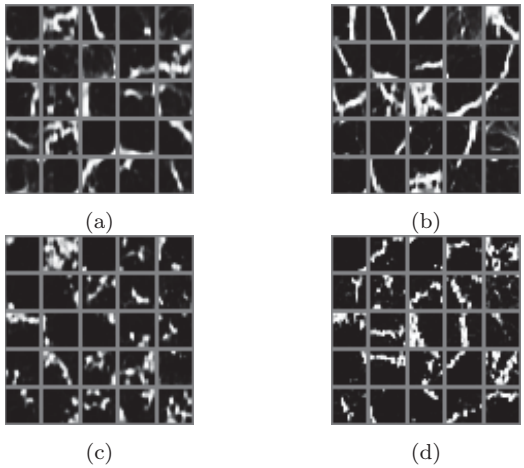


図 5 学習後の RBM および DBM が生成する画像（パッチ）．(a) 隠れノード数 100 の RBM．(b) 隠れノード数 360 の RBM．(c) 各隠れ層に 75 のノードをもつ DBM．(d) 各隠れ層に 200 のノードをもつ DBM．

3.2 実験結果

学習後の RBM と DBM について，各隠れ層のノードが取り出す特徴を図 4 に示す．また，学習された各モデルより生成した画像を図 5 に示す．図 4 より，RBM では 1 層で捉えていたのと類似の特徴を，DBM では 2 層目で捉えているのがわかる．また，図 5 を図 3 と見比べると，学習サンプルに近い画像を生成できていることが分かる．

次に，学習後の RBM と DBM の各モデルを用いて線画の復元を行った．図 6 がその結果である．また，復元精度を 2 画像間のハミング距離によってを評価したものを表 2 に示す．

表 2 から，全体の傾向としてパラメータ数に比例して復元精度が向上していることがまず読み取れる．また RBM と DBM を比較すると，ドットおよび線による比較的軽微

表 1 実験で用いた各モデルの詳細．RBM は隠れ層のノードの数を，DBM は各隠れ層のノード数（2 層で共通）を表す．

パラメータ規模	25000		35000		64000		92000	
	ノードの数	実際のパラメータ数	ノードの数	実際のパラメータ数	ノードの数	実際のパラメータ数	ノードの数	実際のパラメータ数
RBM	100	25956	140	36236	250	64506	360	92772
DBM	75	25231	100	36056	155	64271	200	91856

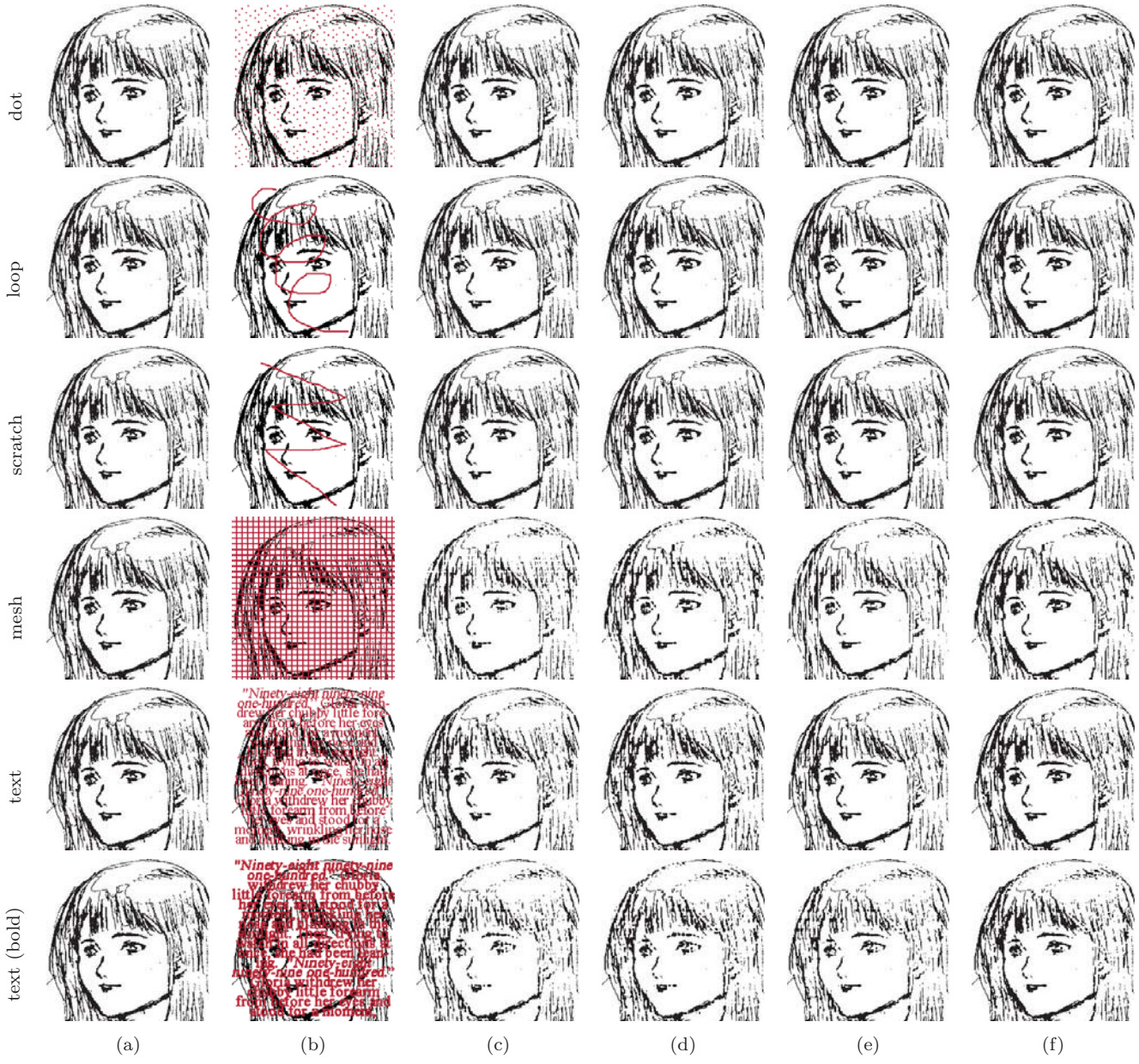


図 6 2 つの規模の RBM および DBM による欠損復元の結果．行ごとに同じ欠損パタンに対する復元結果を示す．各列左から，(a) 元画像，(b) 欠損画像，(c) パラメータ数 25000 の RBM，(d) パラメータ数 25000 の DBM，(e) パラメータ数 92000 の RBM，(f) パラメータ数 92000 の DBM，それぞれの復元結果である．

な欠損に対しては，両者の復元結果に大きな差は見られないが，格子や文字による欠損の場合，DBM の復元結果は RBM のそれを上回っていることが分かる．

図 6 に示した格子による欠損の復元結果をみると，実際，規模の小さい RBM による復元結果では，格子状に線が抜けており，復元がうまく行えていないのがわかる．規模の大きい DBM では，格子状の周期的な復元失敗はほぼ確認できず，画像の復元は概ね成功している．

表 2　元画像と各手法での再構成画像との間のハミング距離

パラメータ数	25000		35000		64000		92000	
	RBM	DBM	RBM	DBM	RBM	DBM	RBM	DBM
dot	235	244	218	223	221	223	202	214
loop	191	171	189	155	163	160	148	147
scratch	134	118	123	113	111	107	94	92
mesh	3625	3225	3461	3044	3246	3136	3382	2893
word	1983	1824	1865	1663	1711	1724	1708	1536
word(bold)	3695	3294	3501	3100	3250	3151	3481	2649

表 2 を詳細に見ると，多層の DBM はパラメータの数倍以上の RBM と同等以上の高い復元精度を達成していることがわかる．例えば格子状の欠損では，パラメータ数が 35000 の DBM による復元結果のハミング距離が 3044 であるのに対して，パラメータ数が 92000 の RBM のそれは 3382 である．同様に文字による欠損では，パラメータ数が 35000 の DBM による復元精度のほうが，パラメータ数 92000 の RBM の復元精度に比べ高精度となっている．

以上より，線画の画像復元の問題では，多層性は確かに効果があり，多層構造をとることでパラメータ数を増やすのと同様にモデルの持つ復元性能を向上させる事ができると分かった．

4. まとめ

本研究では，DBM を用いた線画の無教師学習および，それを利用した線画の欠損復元の方法を示した．実験により，線画の欠損を十分な精度で行えることを示した．また，単層の RBM と多層の DBM の復元精度をネットワークの自由度を揃えて比較することで，多層のネットワークが確かに効果があることを確認した．

参考文献

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, pp. 1106–1114. 2012.

[2] Jurgen Schmidhuber. Multi-column deep neural networks for image classification. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, Conference on Computer Vision and Pattern Recognition '12, pp. 3642–3649, Washington, DC, USA, 2012. IEEE Computer Society.

[3] Geoffrey Hinton and Ruslan Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504 – 507, 2006.

[4] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition. *Signal Processing Magazine*, 2012.

[5] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, pp. 1–8, 2009.

[6] Quoc Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg Corrado, Jeff Dean, and Andrew Ng. Building high-level features using large scale unsupervised learning. In *International Conference in Machine Learning*, 2012.

[7] Honglak Lee, C. Ekanadham, and Andrew Ng. Sparse deep belief net model for visual area V2. *Advances in Neural Information Processing Systems*, pp. 1–8, 2008.

[8] Adam Coates, Andrew Y. Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. *Journal of Machine Learning Research - Proceedings Track*, Vol. 15, pp. 215–223, 2011.

[9] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. *Proceedings of the international ...*, No. 2, 2009.

[10] Junyuan Xie, Linli Xu, and Enhong Chen. Image Denoising and Inpainting with Deep Neural Networks. *... in Neural Information Processing Systems 25*, pp. 1–9, 2012.

[11] S. Kullback and R. A. Leibler. On Information and Sufficiecy. *The Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 79–86, 1951.

[12] A. Miguel and Geoffrey Hinton. On contrastive divergence learning. *Artificial Intelligence and ...*, No. x, 2005.

[13] Tijmen Tieleman. Training restricted Boltzmann machines using approximations to the likelihood gradient. *Proceedings of the 25th international conference on Machine learning - ICML '08*, pp. 1064–1071, 2008.

[14] Bruno A. Olshausen and David J. Fieldt. Sparse coding with an overcomplete basis set: a strategy employed by v1. *Vision Research*, Vol. 37, pp. 3311–3325, 1997.

[15] Ruslan Salakhutdinov and Hugo Larochelle. Efficient learning of deep boltzmann machines. *International Conference on Artificial ...*, Vol. 9, , 2010.

[16] Ruslan Salakhutdinov and Geoffrey Hinton. A Better Way to Pretrain Deep Boltzmann Machines. *Advances in Neural Information Processing Systems*, No. 3, pp. 1–9, 2012.