

cvpaper.challenge

Self-supervised Learning による特徴表現学習

鈴木 智之 ([@tomoyukun](https://twitter.com/tomoyukun))

CVPR 2018 完全読破チャレンジ報告会 cvpaper.challenge勉強会
@Wantedly白金台オフィス

http://hirokatsukataoka.net/project/cc/index_cvpaperchallenge.html

自己紹介

■ 鈴木 智之 (すずき ともゆき)

➤ Twitter : @tomoyukun

➤ 所属 : 慶応大 修士2年

- 青木研究室

- 産総研RA (2017/5~)

- cvpaper.challenge (2017/5~)

➤ 研究の興味

- 行動認識, 表現学習など

➤ 国際発表論文

- Anticipating Traffic Accidents with Adaptive Loss and Large-scale Incident DB, CVPR 2018.

- Learning Spatiotemporal 3D Convolution with Video Order Self-supervision, ECCVWS 2018.

- Semantic Change Detection, ICARCV 2018.



本日の内容

- 教師なし特徴表現学習とは？
- 論文紹介
 - ~ CVPR 2017
 - ~ CVPR 2018
 - さらに最新の動向
- まとめ

教師なし特徴表現学習とは？

教師がないデータを用いてそれらの良い特徴表現を獲得すること（そのまま）

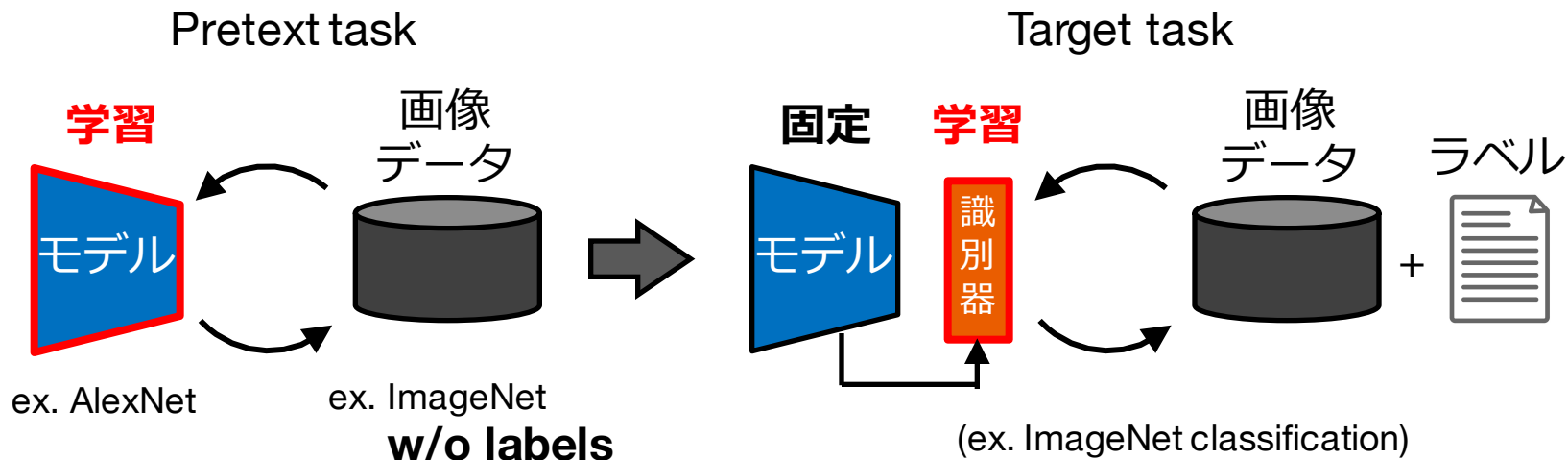
- 今回の**良い**特徴表現 = **discriminative**な特徴表現
 - 解きたいタスク(**target task**)に有効なデータの特徴表現を擬似的なタスク(**pretext task**)を事前に解くことで獲得する
 - disentangle（解釈可能な）など、他の**良さ**については問わない
- Self-supervised Learning
 - **自動で生成できる教師信号**を用いてpretext taskを定義
 - 画像，動画，言語，マルチモーダル
- Self-supervised以外 (Unsupervised)
 - データ分布を表現するモデルを学習する (教師はない)

主に、**画像のみ**を用いた**Self-supervised** Learningについて
(必要に応じてSelf-supervised以外にも触れます)

どうやって良い特徴表現かを評価する？

■ 評価方法①：特徴抽出＋識別器

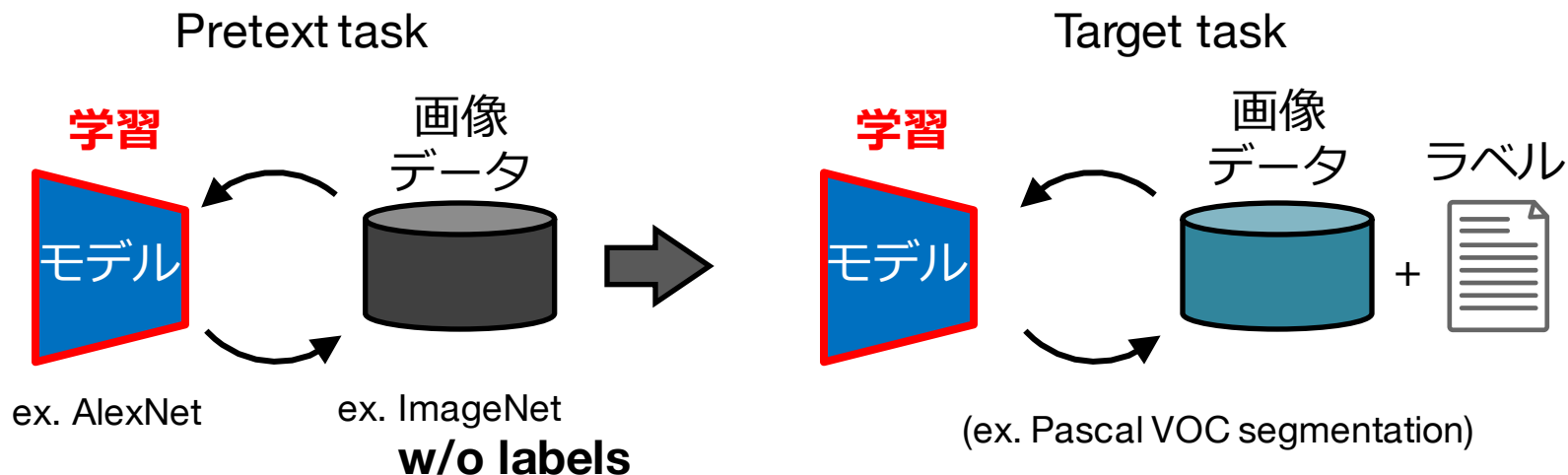
- ▶ Pretext taskで学習したモデルを**重み固定の特徴抽出器**として使い、Target taskでの性能を測る
- ▶ 同じデータセット内で評価することが多い
 - Pretext: **ラベルなし**ImageNet => Target: ラベルありImageNet
- ▶ AlexNetで評価するのがスタンダード (になっってしまった)



どうやって良い特徴表現かを評価する？

■ 評価方法② : Fine-tuning

- ▶ Pretext taskで学習したパラメータを**初期値**として用い, target taskでFine-tuningした時の性能を測る
- ▶ 異なるデータセット間で評価を行うことが多い
 - Pretext : **ラベルなし**ImageNet => Target : ラベルありPascal VOC
- ▶ AlexNetで評価するのがスタンダードなのは評価方法①と同様



今回はラベルなしImageNet => Pascal VOC(cls, det, seg)を基準

Pretext taskの大別

- CVPR2018までの研究をPretext taskを元に大別
- 便宜上の分類であることに注意
 - アイデアベースの手法が多い為、分類が難しい

識別系

Context prediction

Jigsaw

Jigsaw++

Rotation

再構成系

Autoencoder系

Colorization

Split-brain

Context Encoder

Spot Artifact

生成モデル系

VAE系

GAN系

その他

Counting

Noise as target

Instance
Discrimination

Pretext taskの大別

■ 識別系

- 教師なしデータ x に対応する, 自動で得られるカテゴリ t を定義
 - 教師ありデータ $\{x, t\}$ となる
 - x に施された何らかの処理 $\phi(\cdot)$ に応じて t を定義する場合が多い
 - その場合は教師ありデータ $\{\phi(x), t\}$

識別系

Context prediction

Jigsaw

Jigsaw++

Rotation

再構成系

Autoencoder系

Colorization

Split-brain

Context Encoder

Spot Artifact

生成モデル系

VAE系

GAN系

その他

Counting

Noise as target

Instance
Discrimination

Pretext taskの大別

■ 再構成系

- $x = \{x_1, x_2\}$ の一部を観測できている状態で x または x_2 を推定
 - 全て観測できている場合がAuto encoder
 - 回帰学習や条件付き生成モデル的アプローチがある

識別系

Context prediction

Jigsaw

Jigsaw++

Rotation

再構成系

Autoencoder系

Context Encoder

Colorization

Split-brain

Spot Artifact

生成モデル系

VAE系

GAN系

その他

Counting

Noise as target

Instance
Discrimination

Pretext taskの大別

■ 生成モデル系 **基本的にself-supervisedと言われない**

- データ分布 $p(x)$ を学習することに付随して表現を獲得
 - VAEは潜在変数, GANはdiscriminatorの中間特徴など
 - (個人的には) うまく学習できれば一番良い表現を獲得できそう
 - しかし, $p(x)$ の学習が難しい(下界の最大化, ミニマックス問題)

識別系

Context prediction

Jigsaw

Jigsaw++

Rotation

再構成系

Autoencoder系

Colorization

Split-brain

Context Encoder

Spot Artifact

生成モデル系

VAE系

GAN系

その他

Counting

Noise as target

Instance
Discrimination

~ CVPR 2017

Pretext taskの大別

- CVPR2018までの研究をPretext taskを元に大別
- 便宜上の分類であることに注意
 - アイデアベースの手法が多い為、分類が難しい

識別系

Context prediction

Jigsaw

再構成系

AutoEncoder系

Context Encoder

Colorization

Split-brain

生成モデル系

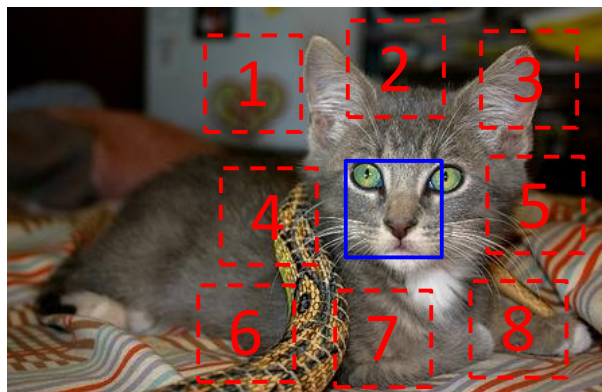
VAE系

GAN系

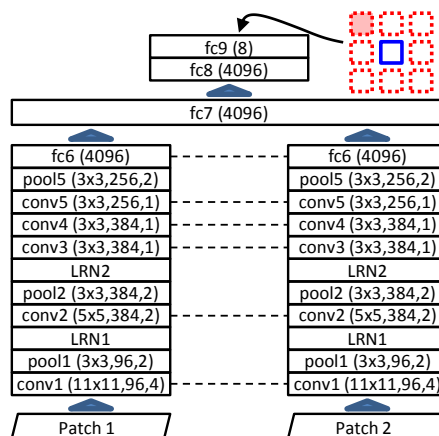
その他

Context Prediction (CP)

- ▶ Pretext task : 画像を3×3に分割し, 二つのパッチの相対位置の8クラス分類
 - 枝構造を持つSiameseNetに2つのパッチを入力
- ▶ 画像のデータ構造を利用した最初のSelf-supervisedな表現学習手法
- ▶ Fine-tuningの結果はランダム初期化より少し良い程度



$$X = \left(\begin{array}{c} \text{cat face} \\ \text{cat ear} \end{array} \right); Y = 3$$



SiameseNet

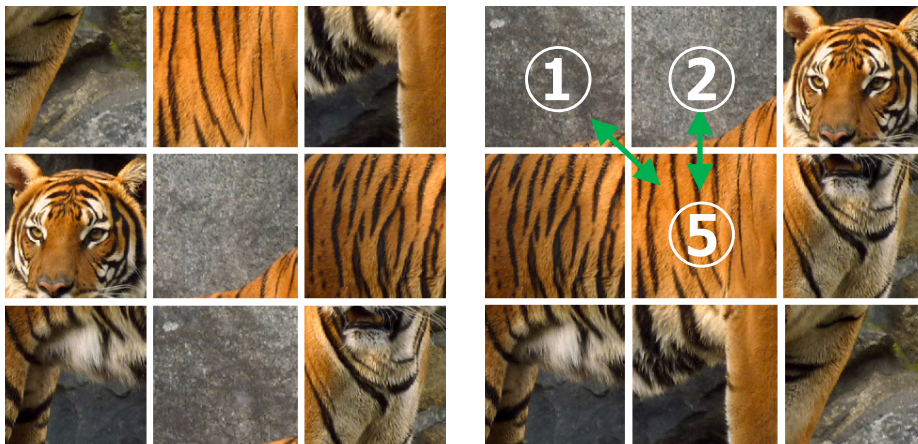
Fine-tuning on Pascal VOC

| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|------|
| random | 53.3 | 43.4 | 19.8 |
| CP | 55.3 | 46.6 | — |

Doersch et al., “Unsupervised visual representation learning by context prediction”, ICCV 2015.

■ Jigsaw Puzzle (JP)

- ▶ Pretext task : パッチをランダムな順に入力し, 正しい順列をクラス識別
 - SiameseNetに9つのパッチを同時に入力
 - 順列は膨大な数になるのでハミング距離が大きくなるように選んだ1000クラスで学習
- ▶ CPはパッチによってはかなりあいまい性がある(下図)
- ▶ **ネットワークが見れるパッチが多い方があいまい性が減る**
- ▶ CPと比較するとかなり精度が改善している



左上や真ん中上の中心からの
相対位置を推定するのはかなり難しい

| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|------|
| random | 53.3 | 43.4 | 19.8 |
| CP | 55.3 | 46.6 | — |
| JP | 67.7 | 53.2 | — |

trivial solution

■ 高次な情報を必要としないPretext taskの解法

- ▶ しかし、実際に捉えてほしいのは高次(semantic)な情報

例えば...

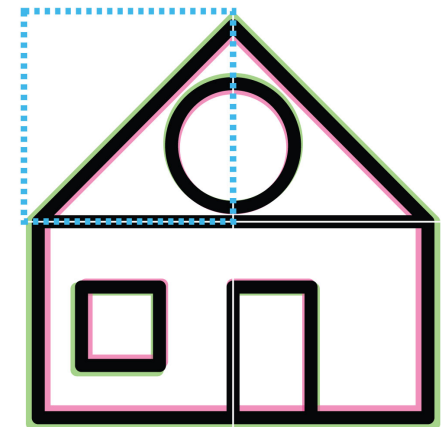
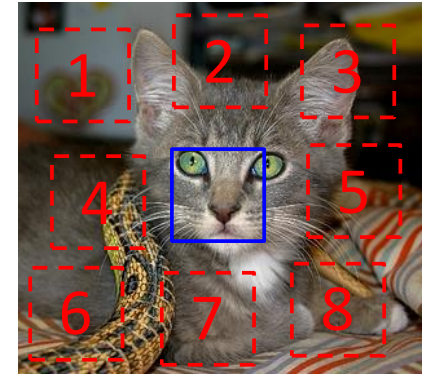
- ▶ パッチ境界の低レベルな情報のみで
相対位置の推定が可能？

- パッチ間にGAPをつける
- パッチ位置をjittering

➡ 境界やその外挿で判断できなく
する

- ▶ 色収差によって相対位置の推定が可能？
 - ランダムに2チャンネルをGaussian noise
に置き換え

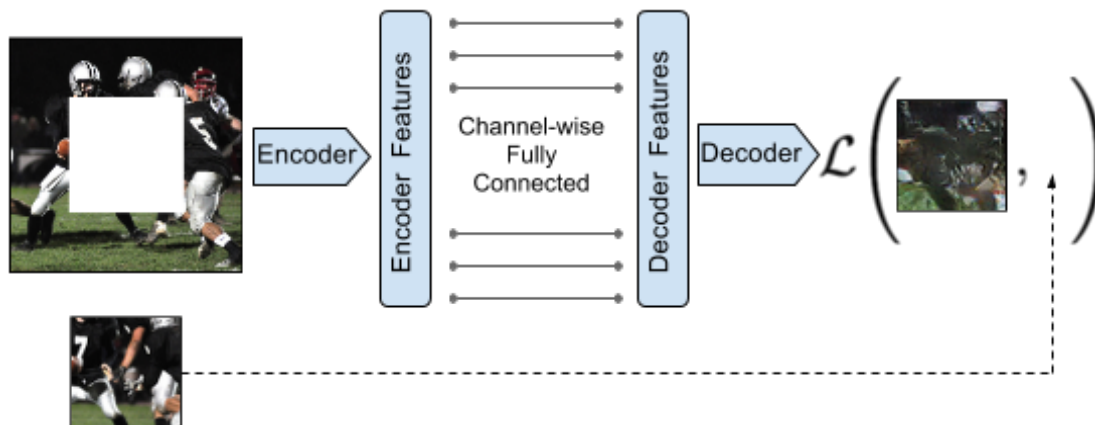
➡ 学習時にチャンネル間の「収差」を
得られなくする



色収差の例

■ Context Encoder (CE)

- Pretext task : 欠損画像の補完
 - Adversarial Loss + L2 Lossを提案しているが, 表現学習の実験は L2 Lossのみ
 - つまりただの回帰
- ネットワークは表現学習の段階で欠損画像しか見ていない

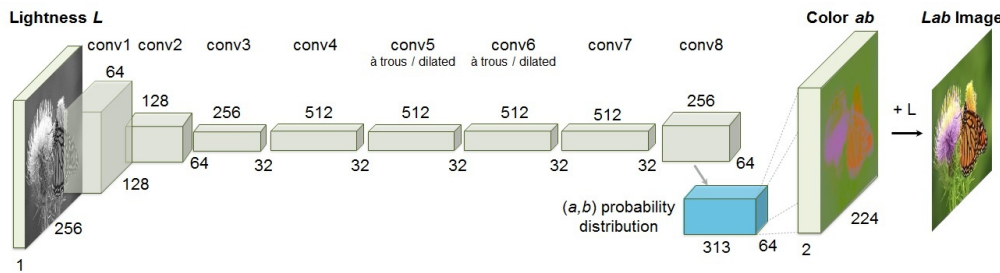


| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|------|
| random | 53.3 | 43.4 | 19.8 |
| CE | 56.5 | 44.5 | 29.7 |
| JP | 67.7 | 53.2 | — |

Pathak et al., "Context encoders: Feature learning by inpainting", CVPR 2016.

Colorful Image Colorization (CC)

- ▶ Pretext task : グレースケール画像の色付け $\{L \Rightarrow ab\}$
- ▶ 単純な回帰ではなく, 量子化したab空間の識別問題を解く
- ▶ グレースケール画像入力を前提として表現学習するため, カラー画像を扱う場合は, Lab入力とし, abチャンネルはランダムに初期化

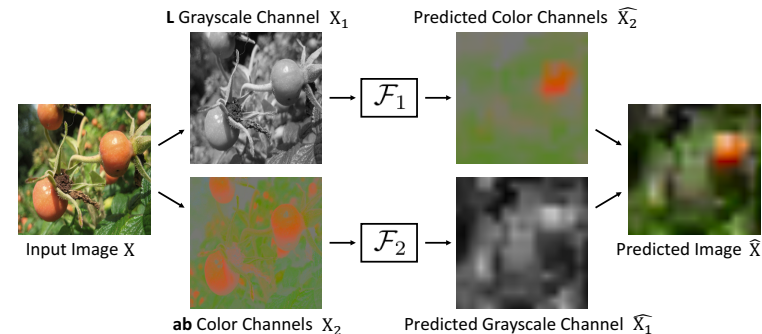


Zhang et al., "Colorful Image Colorization", ECCV 2016.

| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|------|
| random | 53.3 | 43.4 | 19.8 |
| CC | 65.9 | 46.9 | 35.6 |
| SB | 67.1 | 46.7 | 36.0 |
| JP | 67.7 | 53.2 | — |

Split-Brain (SB)

- ▶ ネットワークをチャンネル方向に2分割し, $\{L \Rightarrow ab, ab \Rightarrow L\}$ のアンサンブル
- ▶ **回帰ではなく量子化して識別問題にする方が良い特徴表現が得られた**



Zhang et al., "Split-brain autoencoders: Unsupervised learning by cross-channel prediction", CVPR 2017.

~ CVPR 2018

Pretext taskの大別

- CVPR2018までの研究をPretext taskを元に大別
- 便宜上の分類であることに注意
 - アイデアベースの手法が多い為、分類が難しい

識別系

Context prediction

Jigsaw

Jigsaw++

Rotation

再構成系

AutoEncoder系

Colorization

Split-brain

生成モデル系

VAE系

GAN系

Context Encoder

Spot Artifact

その他

Counting

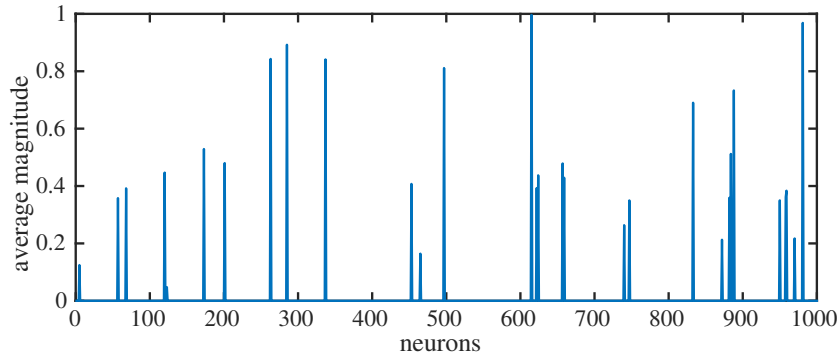
Noise as target

Instance
Discrimination

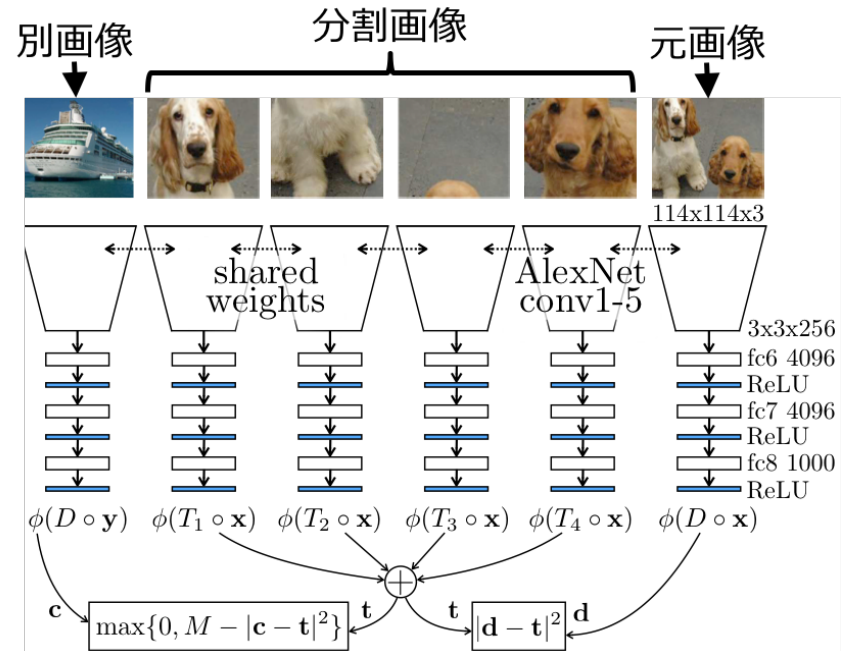
■ Learning to Count (LC)

- ▶ Pretext task : 以下の制約を満たす特徴量を学習
- ▶ 制約 : 各分割画像と元画像をそれぞれ同じCNNに入力し, 元画像の出力特徴が全分割画像の出力特徴の和と一致する
=> 出力特徴の各次元が画像内の「ある高次なprimitive」の量を表す場合に上記の制約を満たすことができる
- ▶ 個人的にかなり面白いアイデア

特徴量がprimitiveのヒストグラムのようなものになる



| | | Cls. | Det. | Seg. |
|-----|--------|-------------|-------------|------|
| | random | 53.3 | 43.4 | 19.8 |
| 同じ人 | LC | 67.7 | 51.4 | 36.6 |
| | JP | 67.7 | 53.2 | — |

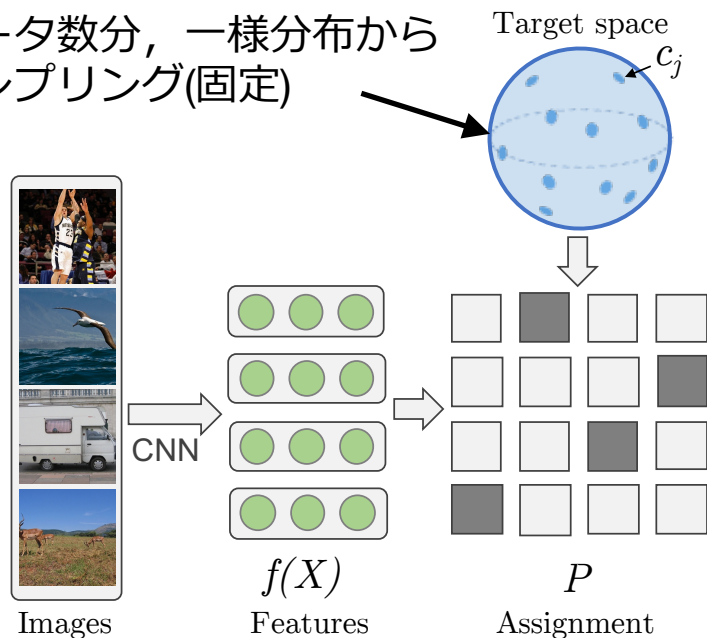


Norozi et al., "Representation Learning by Learning to Count", ICCV 2017.

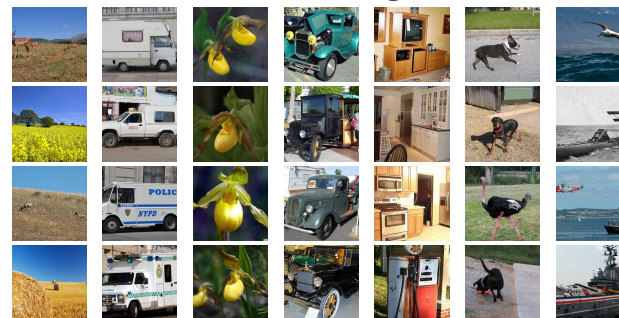
■ Noise as target (NAT)

- Pretext task : 一様にサンプリングされたtarget vectorsに各画像からの出力を**1対1**に対応させ, 近づける
 - Targetは全体サンプルの誤差の和が最小になるように割り当てたい
 - 全走査は厳しいので**バッチごと**にハンガリアン法で近似的に割り当て
- 一見意味不明. 画像の特徴ベクトルを特徴空間上に一様に分散させることに意味があるらしい

データ数分, 一様分布からサンプリング(固定)



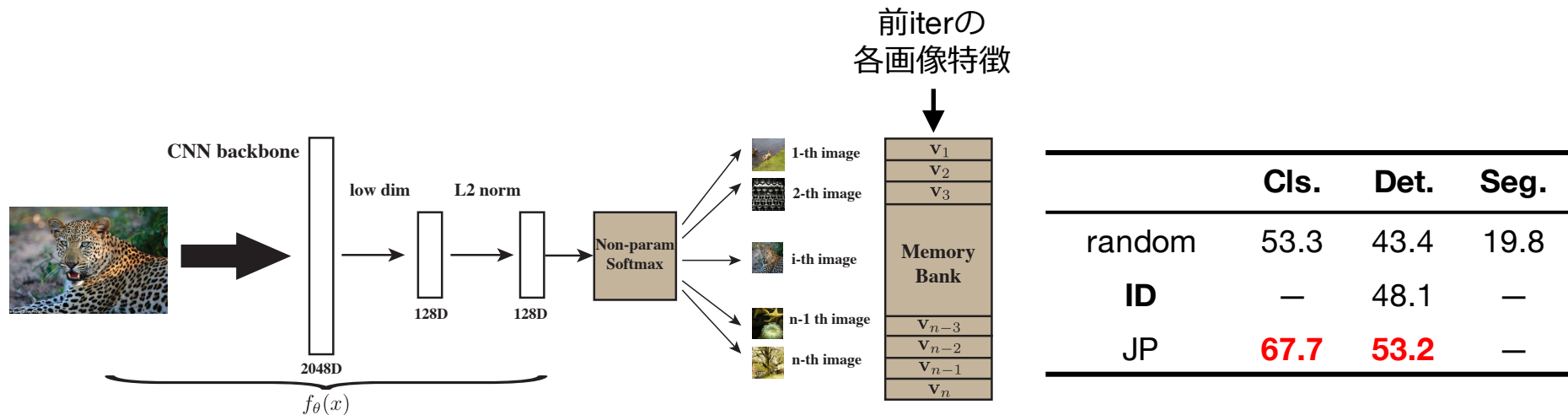
Nearest Neighbor



| | Cls. | Det. | Seg. |
|------------|-------------|-------------|------|
| random | 53.3 | 43.4 | 19.8 |
| NAT | 65.3 | 49.4 | 36.6 |
| JP | 67.7 | 53.2 | — |

■ Instance Discrimination (ID)

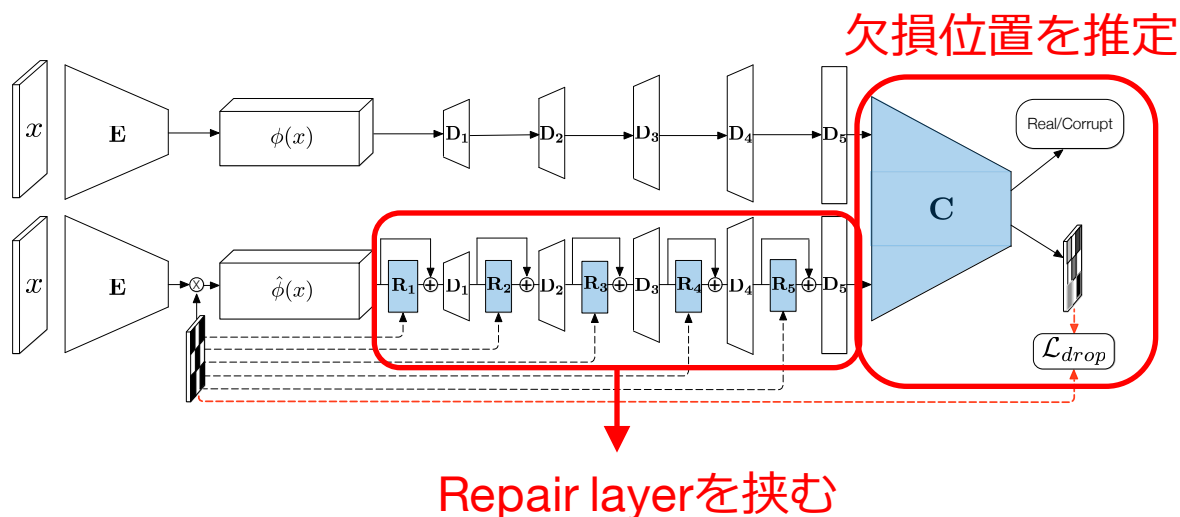
- ▶ Pretext task : 各画像インスタンスを1つのクラスとした識別問題
 - Logitを前iterの各画像特徴と入力画像特徴の内積とした時のCrossEntropyを最小化
 - 実際はクラス数が膨大のため、NCEを用いる
- ▶ 最適な状態としては各画像の特徴ベクトルが超球上にまばらに散るような埋め込みになるはず
=> NATとかなり近いことをしていることになるはず (引用はなし)



Spot Artifact (SA)

- ▶ Pretext task : 特徴マップ上で欠損させた画像の補完
 - 欠損を補完するrepair layersとdiscriminator間で敵対的学習
 - discriminatorが良い特徴表現を得ることを期待
- ▶ 表現学習をどの時点で打ち切るのが難しいはず
 - 敵対的に学習するため, ただロスが下がれば良い特徴を得ているというわけではない
 - 一応, 補完画像のクオリティで評価?

赤 : corrupt, 緑 : real

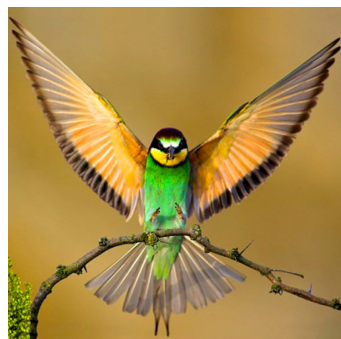


| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|------|
| random | 53.3 | 43.4 | 19.8 |
| SA | 69.8 | 52.5 | 38.1 |
| JP | 67.7 | 53.2 | — |

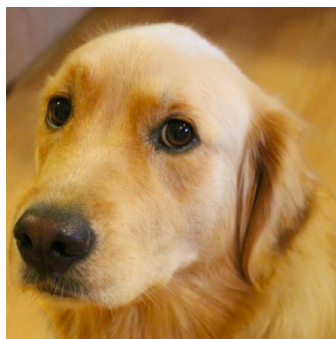
Wu et al., "Self-Supervised Feature Learning by Learning to Spot Artifacts", CVPR 2018.

■ Jigsaw Puzzle++

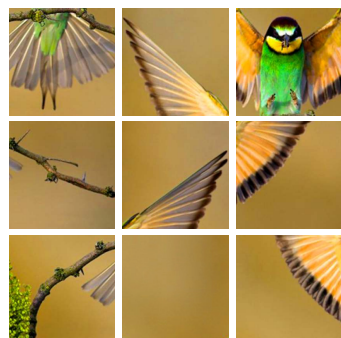
- ▶ Pretext task : 1~3パッチを他の画像のパッチに置き換えたJP
 - **見れる**パッチが少ない・他画像からのパッチを識別する必要がある
 - 複数のクラスに属することがないようにハミング距離を考慮して順列を選択
 - 単純にpretext taskの難度が上がる



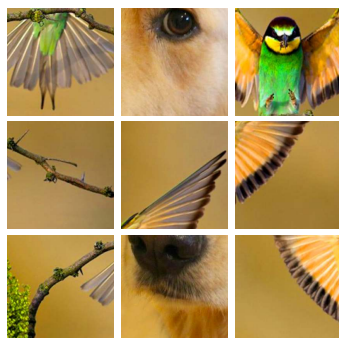
(a)



(b)



(c)



(d)

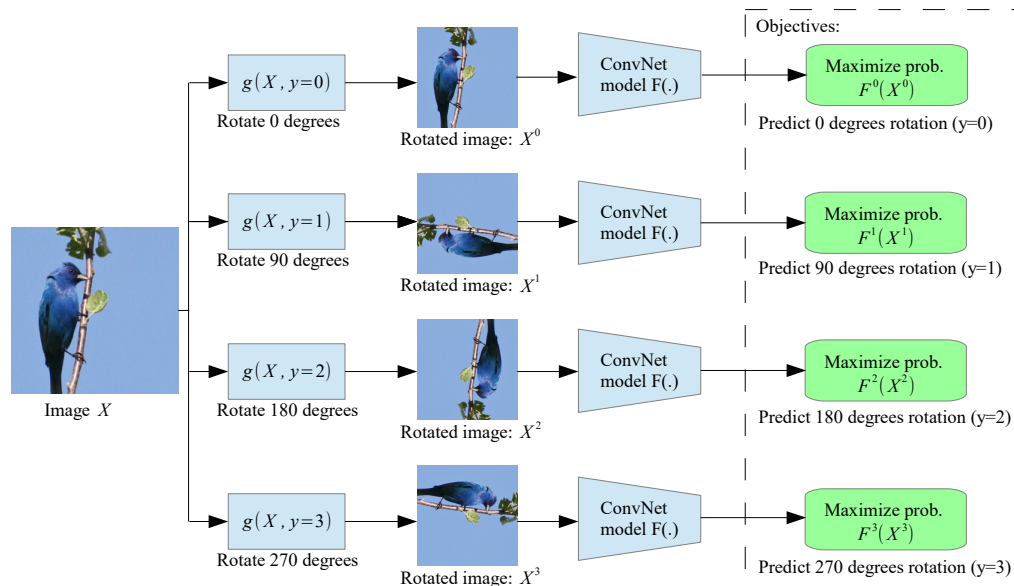
| | Cls. | Det. | Seg. |
|-------------|-------------|-------------|-------------|
| random | 53.3 | 43.4 | 19.8 |
| LC | 67.7 | 51.4 | 36.6 |
| JP++ | 69.8 | 55.5 | 38.1 |
| JP | 67.7 | 53.2 | — |

同じ人 {

Noroozi et al., "Boosting Self-Supervised Learning via Knowledge Transfer", CVPR 2018.

■ Classify Rotation (CR)

- Pretext task : 画像の回転推定
 - $0^\circ, 90^\circ, 180^\circ, 270^\circ$ の4クラス分類
 - それ以上の細かい分類は回転後に補間が必要
=> artifactによるtrivial solutionの原因
- objectの回転角を推定するためにはobjectの高次な情報が必要
- **ここまでの最高精度(Cls., Det.) & 実装が最も簡単**



| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|-------------|
| random | 53.3 | 43.4 | 19.8 |
| CR | 73.0 | 54.4 | 39.1 |
| JP++ | 69.8 | 55.5 | 38.1 |

Gidaris et al., "Unsupervised Representation Learning by predicting Image Rotation", ICLR 2018.

■ Classify Rotation (CR)

- データ構造に依存？
- 画像ドメインによっては低次元特徴で回転の推定が可能では？
 - 実際にPlacesのシーン識別タスクでは奮わない
- 回転が定義できないような画像もあるはず
 - 航空写真など

| Method | Conv1 | Conv2 | Conv3 | Conv4 | Conv5 |
|--|-------------|-------------|-------------|-------------|-------------|
| Places labels Zhou et al. (2014) | 22.1 | 35.1 | 40.2 | 43.3 | 44.6 |
| ImageNet labels | 22.7 | 34.8 | 38.4 | 39.4 | 38.7 |
| Random | 15.7 | 20.3 | 19.8 | 19.1 | 17.5 |
| Random rescaled Krähenbühl et al. (2015) | 21.4 | 26.2 | 27.1 | 26.1 | 24.0 |
| Context (Doersch et al., 2015) | 19.7 | 26.7 | 31.9 | 32.7 | 30.9 |
| Context Encoders (Pathak et al., 2016b) | 18.2 | 23.2 | 23.4 | 21.9 | 18.4 |
| Colorization (Zhang et al., 2016a) | 16.0 | 25.7 | 29.6 | 30.3 | 29.7 |
| Jigsaw Puzzles (Noroozi & Favaro, 2016) | <u>23.0</u> | <u>31.9</u> | 35.0 | 34.2 | 29.3 |
| BIGAN (Donahue et al., 2016) | 22.0 | 28.7 | 31.8 | 31.3 | 29.7 |
| Split-Brain (Zhang et al., 2016b) | 21.3 | 30.7 | 34.0 | 34.1 | 32.5 |
| Counting (Noroozi et al., 2017) | 23.3 | 33.9 | 36.3 | 34.7 | 29.6 |
| (Ours) RotNet | 21.5 | 31.0 | <u>35.1</u> | <u>34.6</u> | 33.7 |

Places



例えば、空の位置のみで
回転推定できる

比較

{Self, Un}-supervised learning on ImageNet => Fine-tuning on Pascal VOC2007

| Method | Conference | Classification (%mAP) | Detection (%mAP) | Segmentation (%mIoU) |
|--------------------|------------|-----------------------|------------------|----------------------|
| Random init. | — | 53.3 | 43.4 | 19.8 |
| Context prediction | ICCV15 | 55.3 | 46.6 | — |
| Context encoder | CVPR16 | 56.5 | 44.5 | 29.7 |
| Colorize | ECCV16 | 65.9 | 46.9 | 35.6 |
| Jigsaw | ECCV16 | 67.7 | 53.2 | — |
| Split-Brain | CVPR17 | 67.1 | 46.7 | 36.0 |
| NAT | ICML17 | 65.3 | 49.4 | 36.6 |
| Counting | ICCV17 | 67.7 | 51.4 | 36.6 |
| Rotation | ICLR18 | 73.0 | 54.4 | 39.1 |
| Spot Artifact | CVPR18 | 69.8 | 52.5 | 38.1 |
| Instance Dis. | CVPR18 | — | 48.1 | — |
| Jigsaw++ | CVPR18 | 69.8 | 55.5 | 38.1 |
| Supervised | — | 79.9 | 59.1 | 48.0 |

比較

{Self, Un}-supervised learning on ImageNet => Fine-tuning on Pascal VOC2007

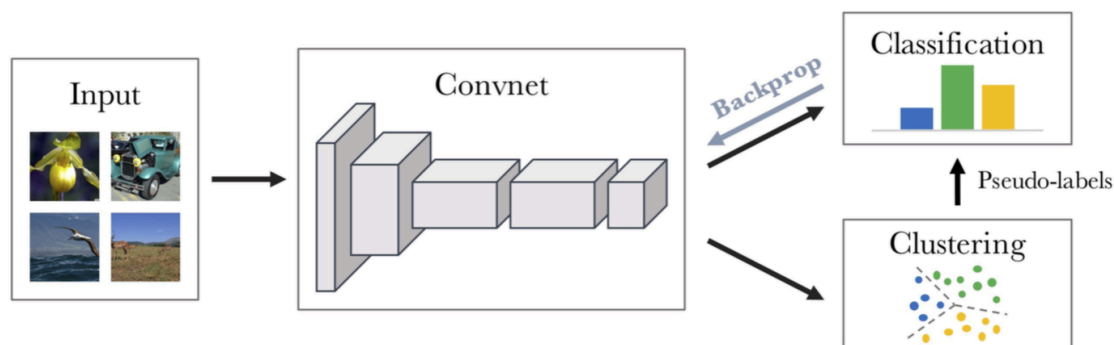
| Method | Conference | Classification (%mAP) | Detection (%mAP) | Segmentation (%mIoU) |
|---------------------|---------------|-----------------------|------------------|----------------------|
| Random init. | — | 53.3 | 43.4 | 19.8 |
| Context prediction | ICCV15 | 55.3 | 46.6 | — |
| Context encoder | CVPR16 | 56.5 | 44.5 | 29.7 |
| Colorize | ECCV16 | 65.9 | 46.9 | 35.6 |
| Jigsaw | ECCV16 | 67.7 | 53.2 | — |
| Split-Brain | CVPR17 | 67.1 | 46.7 | 36.0 |
| NAT | ICML17 | 65.3 | 49.4 | 36.6 |
| Counting | ICCV17 | 67.7 | 51.4 | 36.6 |
| Rotation | ICLR18 | 73.0 | 54.4 | 39.1 |
| Spot Artifact | CVPR18 | 69.8 | 52.5 | 38.1 |
| Instance Dis. | CVPR18 | — | 48.1 | — |
| Jigsaw++ | CVPR18 | 69.8 | 55.5 | 38.1 |
| Deep Cluster | ECCV18 | 73.7 | 55.4 | 45.1 |
| Supervised | — | 79.9 | 59.1 | 48.0 |

さらに最新の動向

最新動向

■ Deep Cluster (DC)

- 以下の操作を繰り返し行う
 - ① CNNの中間特徴を元にk-meansクラスタリング
 - ② 割り当てられたクラスをPseudo labelとして識別問題を学習
- 最初のiterationではランダム初期化されたCNNの出力を元にクラスタリング
 - その出力を用いてMLPを学習しても12%出る => 入力情報はある程度保持されてる
- ImageNetでの実験ではk = 10000 (> 1000)が最も良い
- 単純かつ非常に強力な手法



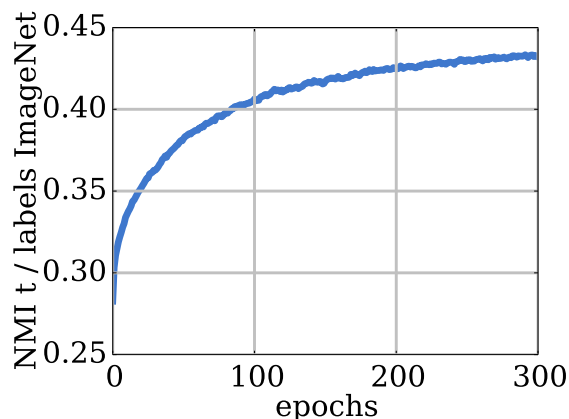
| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|-------------|
| random | 53.3 | 43.4 | 19.8 |
| CR | 73.0 | 54.4 | 39.1 |
| JP++ | 69.8 | 55.5 | 38.1 |
| DC | 73.7 | 55.4 | 45.1 |

最新動向

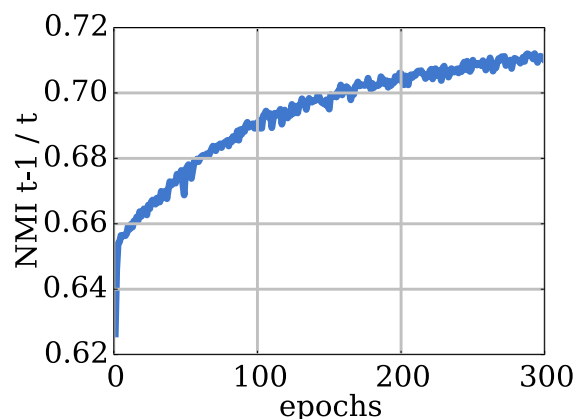
■ Deep Cluster (DC)

- 以下の操作を繰り返し行う
 - ① CNNの中間特徴を元にk-meansクラスタリング
 - ② 割り当てられたクラスタをPseudo labelとして識別問題を学習
- 最初のiterationではランダム初期化されたCNNの出力を元にクラスタリング
 - その出力を用いてMLPを学習しても12%出る => 入力情報はある程度保持されてる
- ImageNetでの実験ではk = 10000 (> 1000)が最も良い
- 単純かつ非常に強力な手法

ImageNet labelとクラスタの
相互情報量が増加していく



epoch間の相互情報量が増加
=> クラスタ割り当てが安定

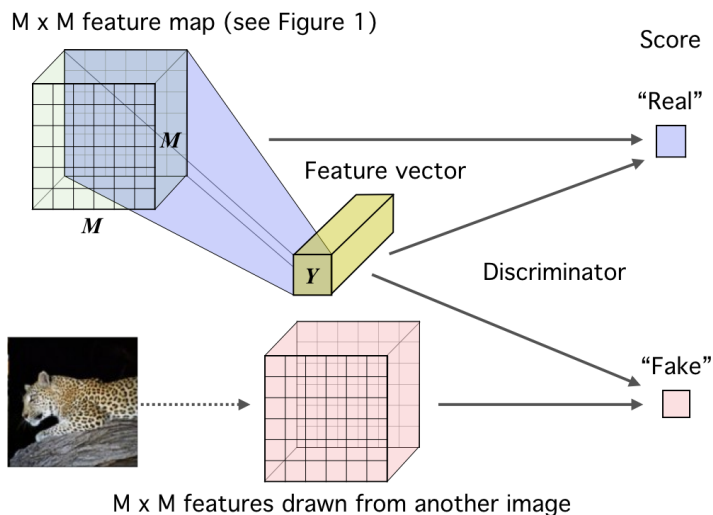


| | Cls. | Det. | Seg. |
|-----------|-------------|-------------|-------------|
| random | 53.3 | 43.4 | 19.8 |
| CR | 73.0 | 54.4 | 39.1 |
| JP++ | 69.8 | 55.5 | 38.1 |
| DC | 73.7 | 55.4 | 45.1 |

最新動向

■ Deep INFORMAX (DIM)

- ▶ 入力 x と特徴ベクトル z の相互情報量 $I(x; z)$ を最大化するように学習
 - 簡単に言うと x と z の依存を大きくする
 - 実際には z と x の**各バッチ**の相互情報量最大化が大きな効果を発揮
- ▶ (x, z) の**positive or negativeペアの識別をするdiscriminatorをつけてend-to-endに学習するだけで $I(x; z)$ の下限を最大化することができる**
- ▶ GANのような交互最適化でもないので, 実装・学習が簡単
- ▶ 全ての手法との比較はしていないが教師あり学習に近い精度



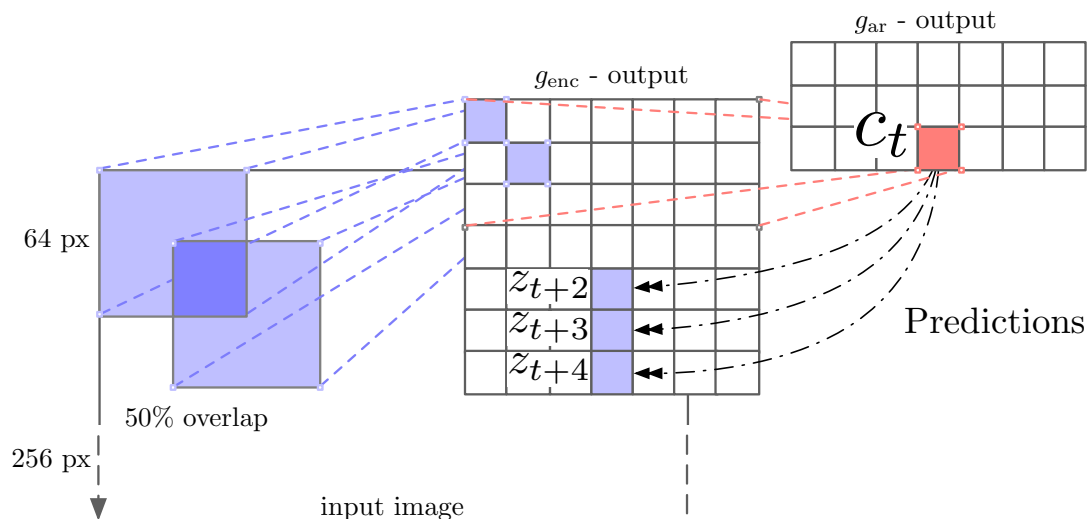
Tiny ImageNetにおいて教師ありに近い精度

| | Tiny ImageNet | | |
|------------------|---------------|-------------|-------------|
| | conv | fc (4096) | Y(64) |
| Fully supervised | | 36.60 | |
| VAE | 18.63 | 16.88 | 11.93 |
| AAE | 18.04 | 17.27 | 11.49 |
| BiGAN | 24.38 | 20.21 | 13.06 |
| NAT | 13.70 | 11.62 | 1.20 |
| DIM(G) | 11.32 | 6.34 | 4.95 |
| DIM(L) | 33.8 | 34.5 | 30.7 |

最新動向

■ Contrastive Predictive Coding (CPC)

- ▶ 系列情報においてある時点での特徴ベクトル c_t と先の入力 x_{t+1} 間の相互情報量を最大化
- ▶ こちらはdiscriminatorがN個のペアから1つのpositiveペアを識別するNクラス分類を解くことで相互情報量の下界を最大化
- ▶ 画像の場合は図のように系列を特徴マップの上から下の方向として捉える
- ▶ 全ての手法との比較はしていないが実験内では圧倒的な精度



| Method | Top-1 ACC |
|----------------------------|-------------|
| Using AlexNet conv5 | |
| Video [27] | 29.8 |
| Relative Position [11] | 30.4 |
| BiGan [34] | 34.8 |
| Colorization [10] | 35.2 |
| Jigsaw [28] * | 38.1 |
| Using ResNet-V2 | |
| Motion Segmentation [35] | 27.6 |
| Exemplar [35] | 31.5 |
| Relative Position [35] | 36.2 |
| Colorization [35] | 39.6 |
| CPC | 48.7 |

まとめ

まとめ

■ CVPR2018まで

- ▶ アイデアベースで多様な手法が発表されてきた（お蔵入もたくさんあったはず）
- ▶ 画像のデータ構造に着目したSelf-supervised learningが精度的にも優位だった（Rotation, Jigsaw...）

■ 現在の動き

- ▶ データ構造に依存しない手法がうまくいきはじめた（Deep Cluster, 相互情報量に着目したアプローチ）
- ▶ データ構造に依存した手法は画像データのドメインによってうまくいくかが左右される考え（rotation on Placesの結果参照）（持論）

■ 今後の展望

- ▶ 手法的な展望
 - データ構造に依存しない手法がさらに発展（具体的には想像がつかない）
 - データ構造に{依存した + 依存しない}手法（Rotation + Deep Cluster）
- ▶ 研究領域としての展望
 - 打倒教師あり学習（ImageNet pretrainedを超える）
 - Task-specificな教師なし学習（現在もありますが...）
こちらの方がデータ構造に着目するself-supervised learningと相性が良さそうに感じる

まとめ

■ 実用として

- ▶ 学習済みモデルとしてはImageNet pretrained modelを使用すれば良い風潮
- ▶ しかし, ImageNet pretrained modelが有効でない場合もある
 - 画像のドメインが大きく異なる場合
- ▶ そういった条件では実用段階
 - 特に実装が簡単なDeep Clusterなど
- ▶ 条件によっては半教師あり学習と競合する場合も
 - 教師なしデータ + 教師ありデータ

ありがとうございました