

深層学習のリモートセンシングへの応用

株式会社パスコ 濱口竜平



PASCO CORPORATION

自己紹介

- 濱口竜平
- ~2016: 東京大学 航空宇宙工学専攻
 - 超小型人工衛星開発
- 2016~: 株式会社パスコ
 - CNNを用いた衛星画像からの建物検出
- 2016~: 産総研人工知能研究センターに出向
 - 2時期画像からの変化検出



本日の内容

- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - 道路検出はトポロジーまで考える
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

なぜ深層学習か？

衛星画像データの質・量のともに増大

- 空間分解能の向上: 商用衛星で30cm (Worldview-3)
- 観測頻度の向上: 1日1回以上 (小型衛星数十機によるコンステレーション)



Skysat (Planet)

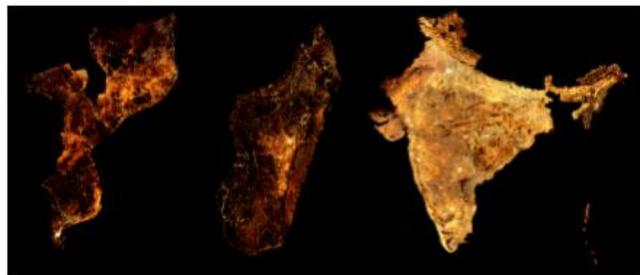
世界のコンステレーション計画 [1]

企業名	撮像頻度 1回/日以上	センサ種類	機数	重量 (kg)
Astro Digital (米)	●	Optical	30	20
Axelspace (日)	●	Optical	50	95
BlackBridge (Planet) (米)	●	Optical	5	150
BlackSky Global (米)	●	Optical	60	50
Capella Space (米)	●	Radar	30	TBD
XpressSAR (米)		Radar	4	TBD
GeoOptics (米)	●	Radio occultation	24	115
HawkEye360 (米)	●	RF mapping	21+	TBD
Hera Systems (米)	●	Optical	48	24
ICEYE (フィンランド)	●	Radar	50	<100
PlanetiQ (米)	●	Radio occultation	12	22
Planetary Resources (米)	●	Optical	10	TBD
Planet (米)	●	Optical	100+	3
Satellite (米)	●	Optical	25+	35
Spire Global (米)		Radio occultation	50	3
Terra Bella (Planet) (米)	●	Optical	24	120

[1] 経済産業省, “コンステレーションビジネス時代の到来を見据えた 小型衛星・小型ロケットの技術戦略に関する研究会 報告書” 平成30年5月

深層学習を使って何ができるか？

人口推定



(a) Mozambique (b) Madagascar (c) India

(Zhang et al., 2017)

地図自動更新



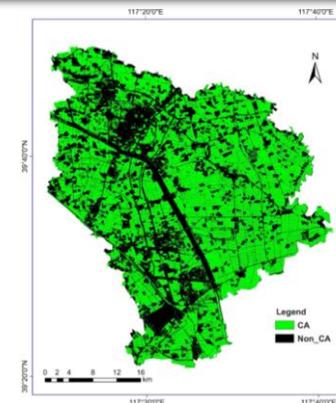
(Batra et al., 2019)

海洋監視



(Yang et al., 2018)

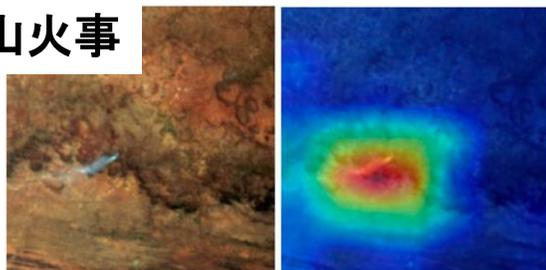
農地管理



(Du et al., 2019)

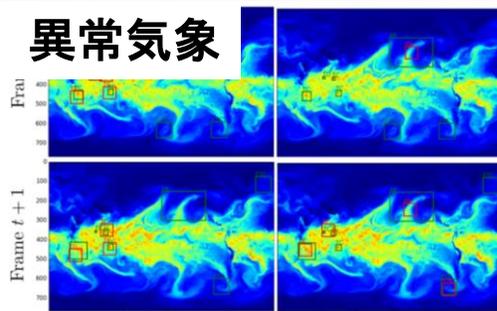
災害監視

山火事



(Ba et al., 2019)

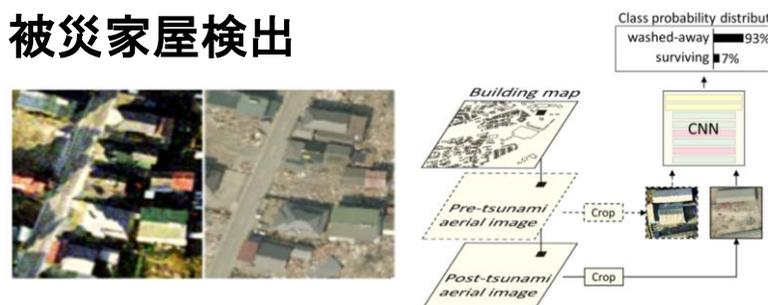
異常気象



(Racah et al., 2017)

被災状況把握

被災家屋検出



(Fujita et al., 2017)

本日の内容

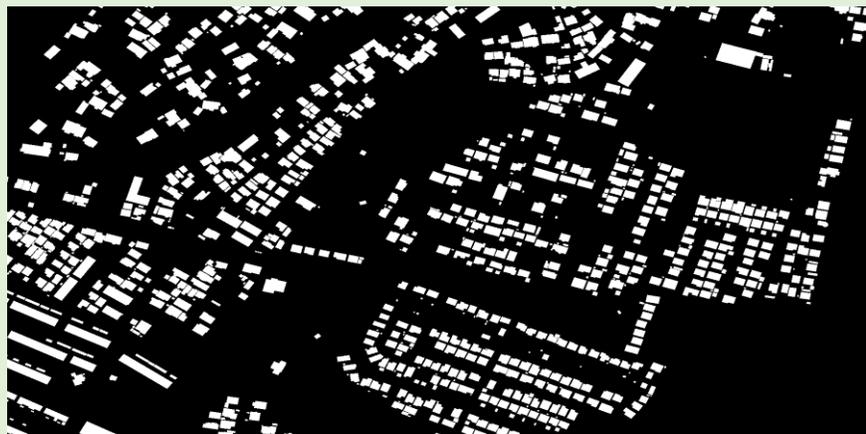
- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - 道路検出はトポロジーまで考える
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

リモートセンシング画像と地上視点画像の違い

航空機画像からの建物検出



1024pixel



車載画像のセグメンテーション



1024pixel



対象とするオブジェクトのサイズが小さく、かつ密集している

Q:地上視点画像とリモートセンシング画像で最適なネットワーク構造は同じか？

Q:地上視点画像とリモートセンシング画像で最適なネットワーク構造は同じか？

A:実は同じではない

Q:地上視点画像とリモートセンシング画像で最適なネットワーク構造は同じか？

A:実は同じではない

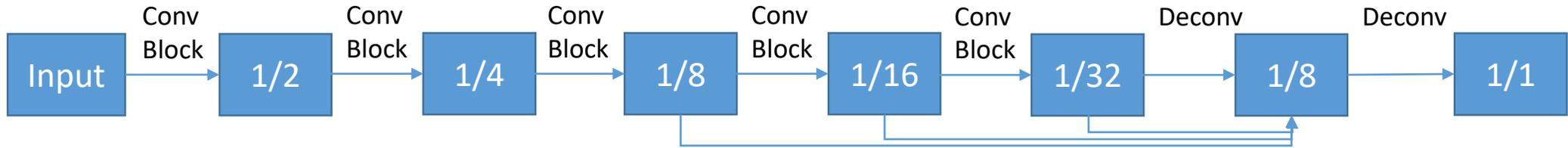


小さなオブジェクトを認識するのに適した構造を設計する必要がある

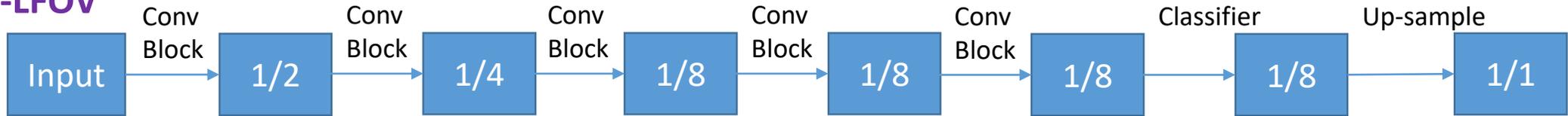
ダウンサンプリングによる解像度の低下

コンピュータビジョンの分野で有名なセグメンテーションモデル

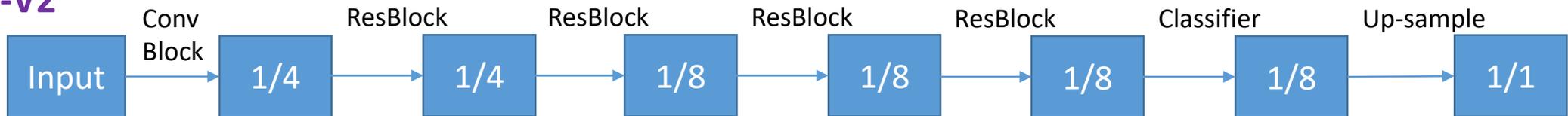
FCN-8s



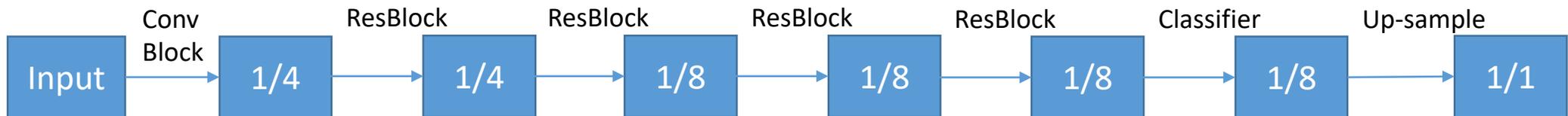
DeepLab-LFOV



DeepLab-V2



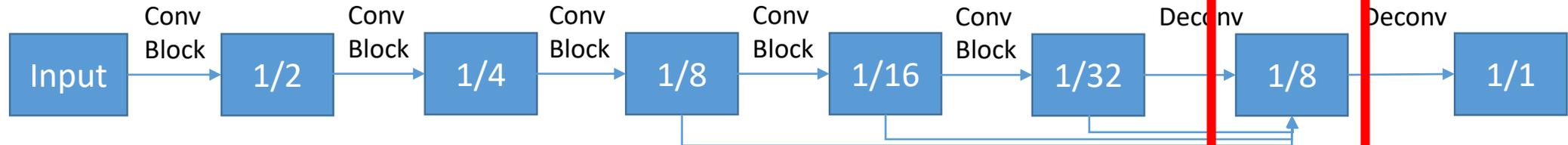
PSPNet



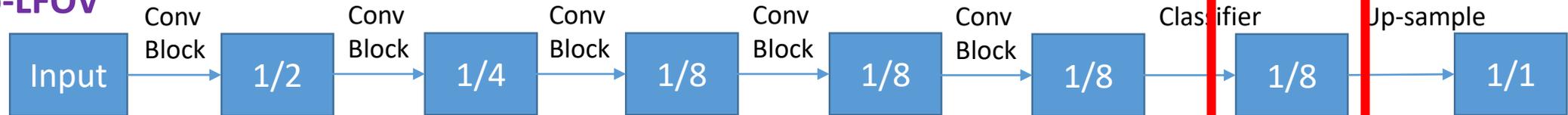
ダウンサンプリングによる解像度の低下

出力を計算するのに用いる情報は1/8まで解像度が落ちている

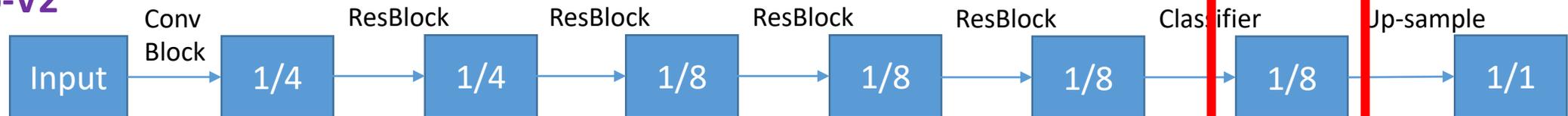
FCN-8s



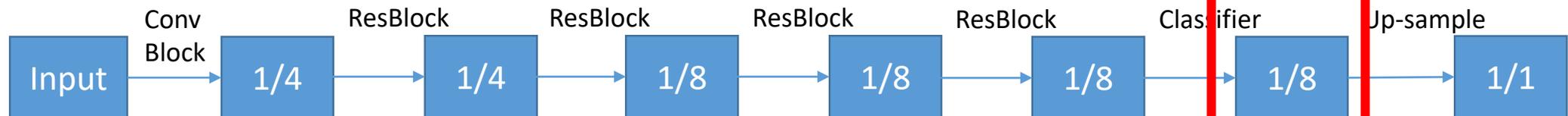
DeepLab-LFOV



DeepLab-V2



PSPNet



ダウンサンプリングによる解像度の低下



FCN-8s



DeepLab-LFOV



DeepLab-V2



解像度の低い特徴を使っているため、建物の境界をうまく捉えることができない

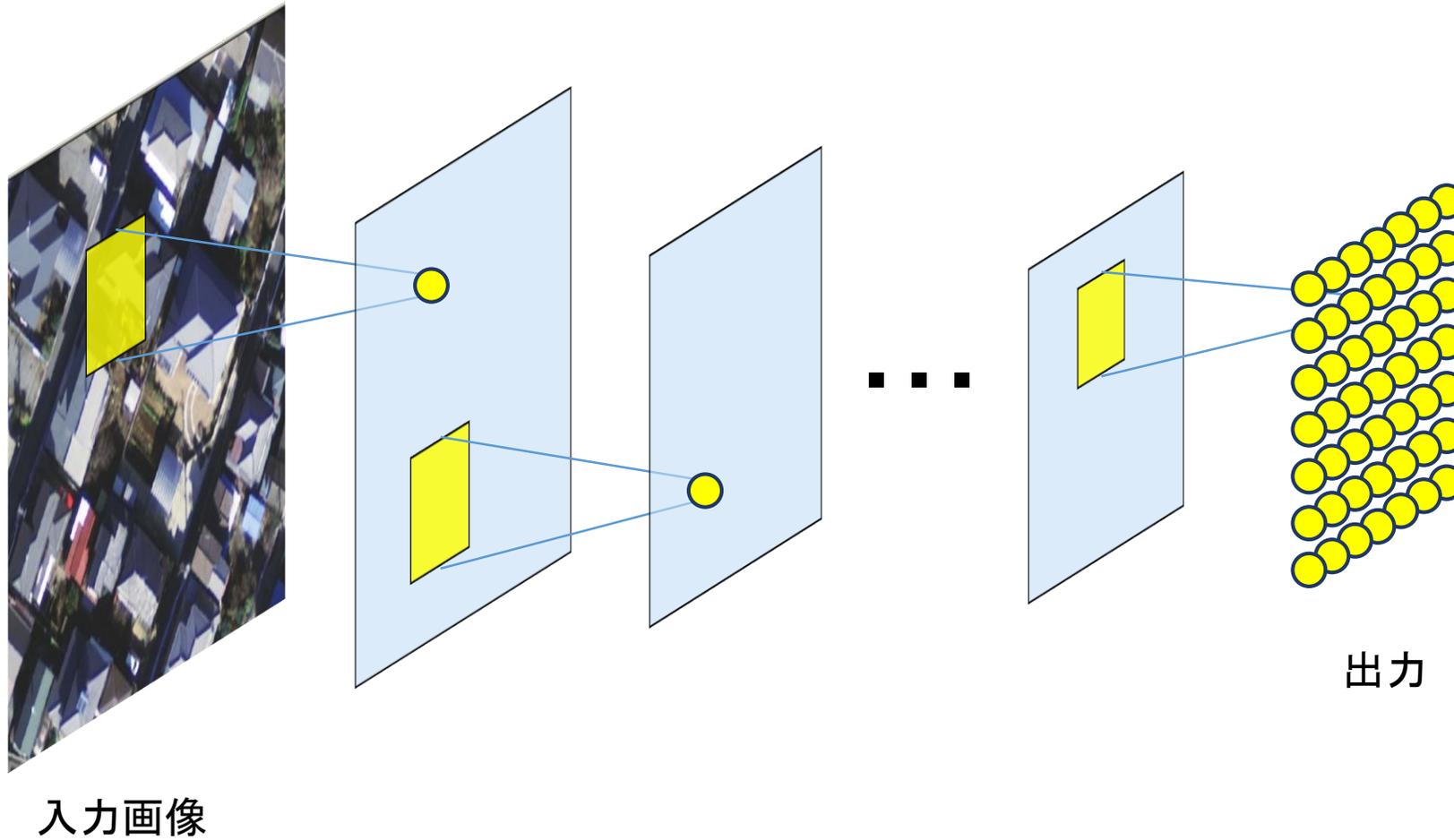
ダウンサンプリングはいらない？

ダウンサンプリングの重要な役割のひとつは“CNNの受容野を広げる”こと

- ダウンサンプリングは空間的に情報を要約していく処理
- 要約することでより広い範囲の情報を取り込むことができる

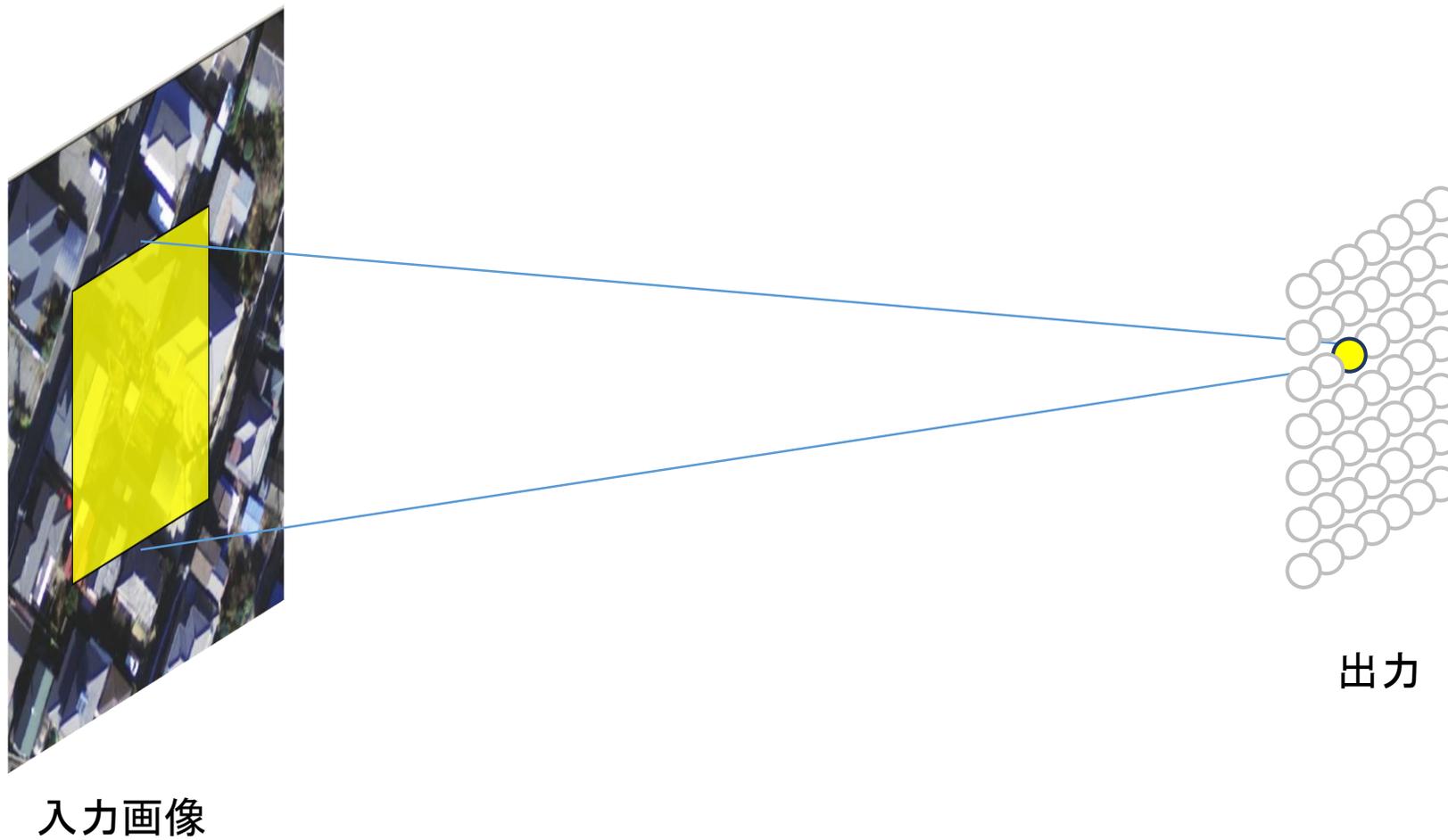
受容野とは

ざっくり言うとニューラルネットの視野。画像のどの範囲の情報を処理しているかを表すもの。



受容野とは

ざっくり言うとニューラルネットの視野。画像のどの範囲の情報を処理しているかを表すもの。



ダウンサンプリングはいらない？

ダウンサンプリングの重要な役割のひとつは“CNNの受容野を広げる”こと

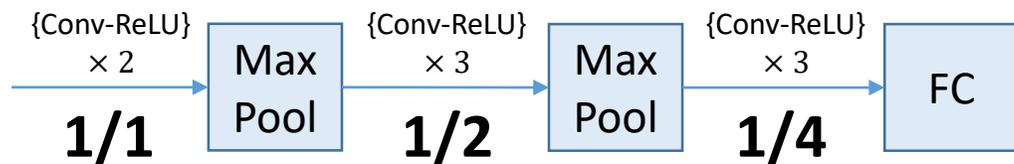
- ダウンサンプリングは空間的に情報を要約していく処理
- 要約することでより広い範囲の情報を取り込むことができる



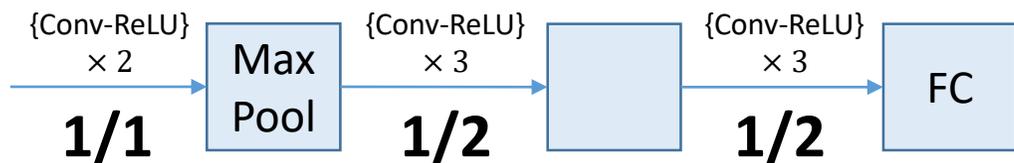
小さいオブジェクトの認識にも広い受容野は必要なのか？
ダウンサンプリングをなくして高い解像度のまま情報処理を行った方が得策ではないか？

実際に確かめてみた

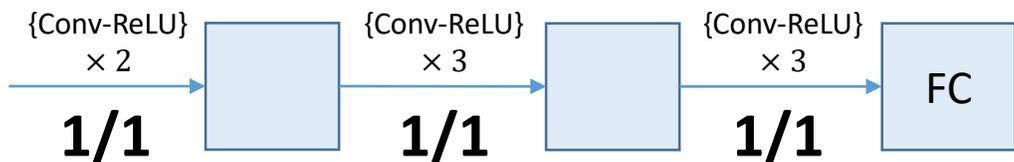
Pool × 2



Pool × 1



Poolなし



タスク: 衛星画像 (50cm 分解能) からの建物検出



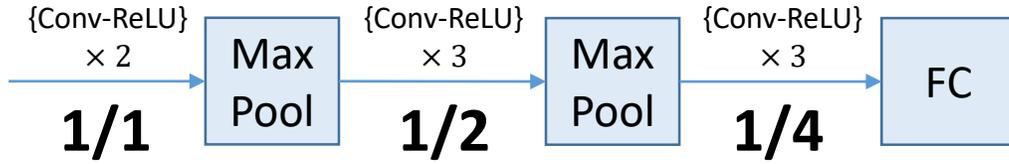
画像

正解

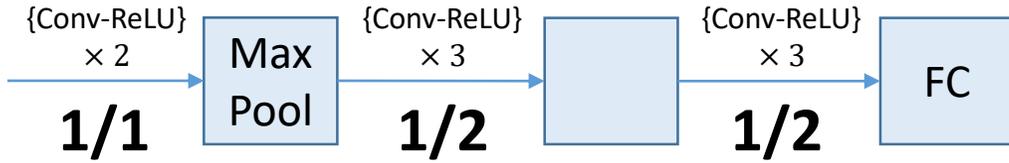
モデル	受容野サイズ	Pixel F1	mIoU
Pool × 2	72x72 pixel	58.59%	67.34%
Pool × 1	42x42 pixel	57.82%	66.89%
Poolなし	23x23 pixel	53.92%	64.59%

実際に確かめてみた

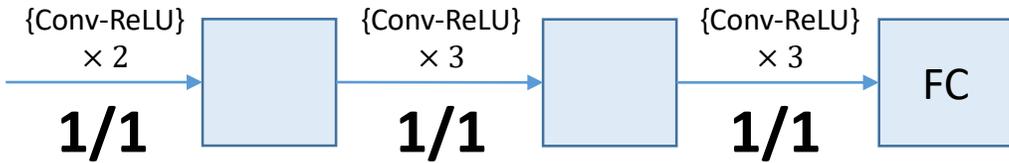
Pool × 2



Pool × 1



Poolなし



タスク: 衛星画像(50cm分解能)からの建物検出



画像

正解

モデル	受容野サイズ	Pixel F1	mIoU
Pool × 2	72x72 pixel	58.59%	67.34%
Pool × 1	42x42 pixel	57.82%	66.89%
Poolなし	23x23 pixel	53.92%	64.59%

小さなオブジェクトの認識に受容野の広さも重要

受容野の重要性

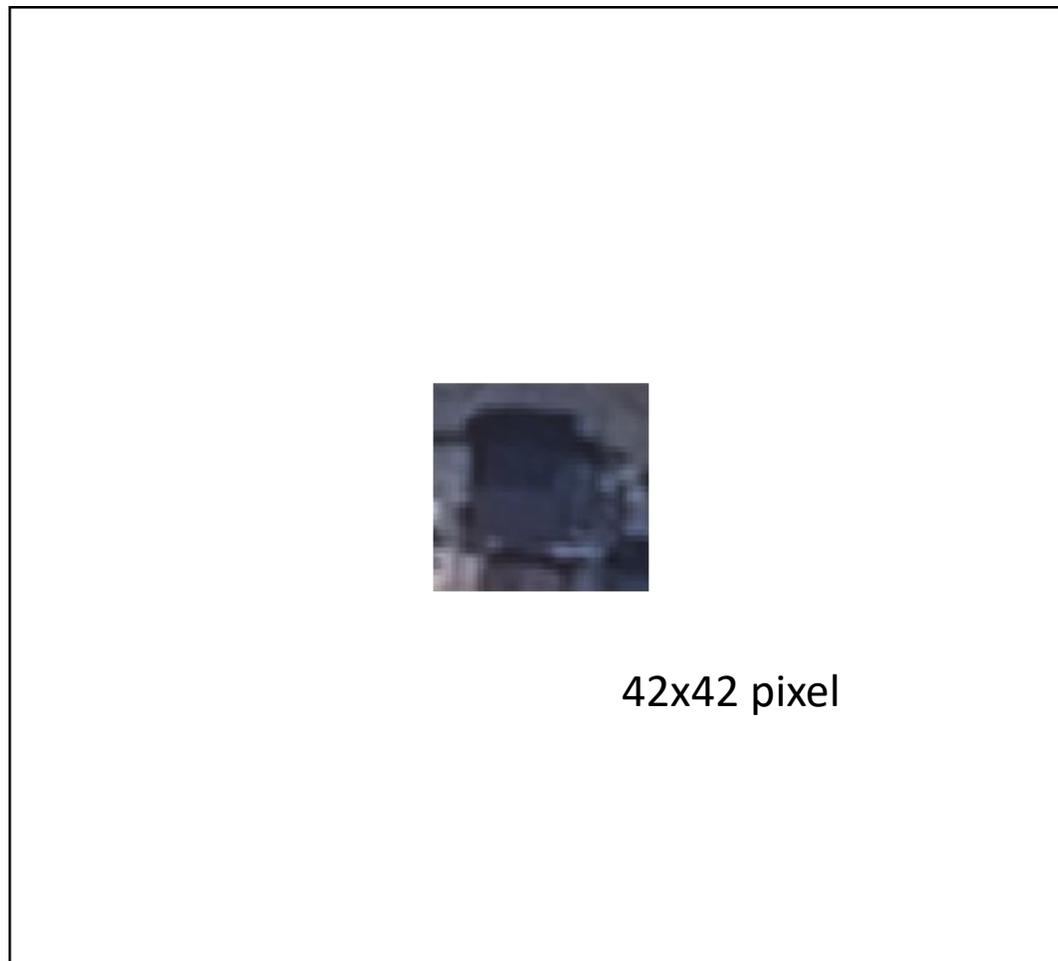


受容野の重要性



“Poolなし”の視野

受容野の重要性



“Pool × 1” の視野

受容野の重要性



72x72 pixel

“Pool × 2” の視野

受容野の重要性



190x190 pixel

まとめると

リモートセンシング画像の認識では・・・

- 解像度を落としたくない
- 受容野も広げたい



これをどう両立するかが課題

まとめると

リモートセンシング画像の認識では・・・

- 解像度を落としたくない
- 受容野も広げたい



これをどう両立するかが課題



Skip connectionを用いる手法

- ✓ 受容野はPoolingで広げる
- ✓ 浅い層の特徴を再利用することで失った解像度を回復



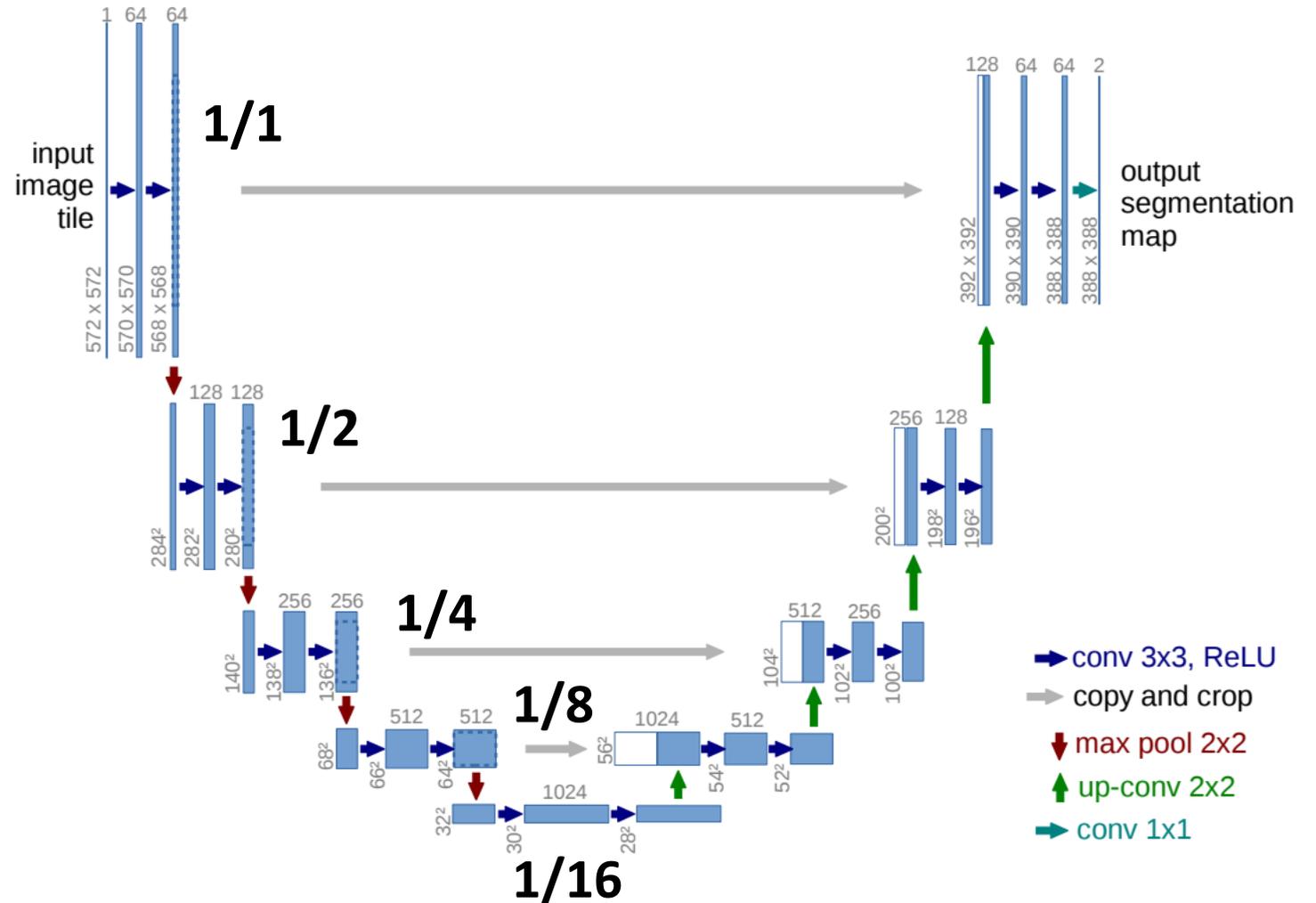
Dilated convolutionを用いる手法

- ✓ 解像度を落とさずに受容野を広げることのできる特殊な畳み込みカーネルを用いる

Skip connection系

- Poolingで失われた解像度を浅い層にあるlow-levelな特徴量を用いて回復する
- Biomedical分野で登場したものだが、Remote Sensingの問題とも親和性が良い
- リモセン系のコンペの上位はたいていUNetの派生形

UNet [Ronneberger et al., 2015]



UNetの効果



FCN-8s



DeepLab-LFOV



DeepLab-V2

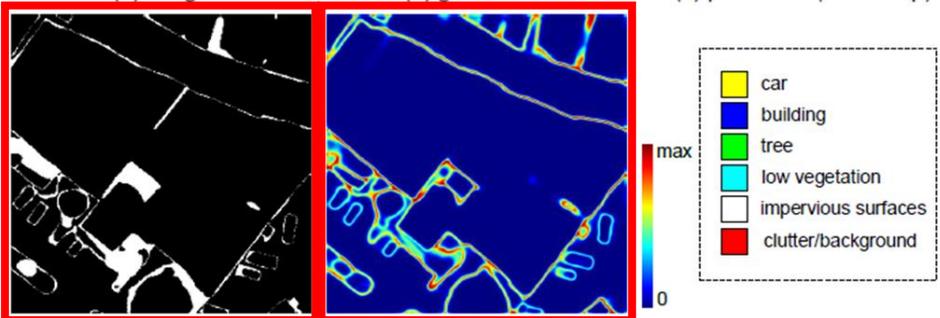
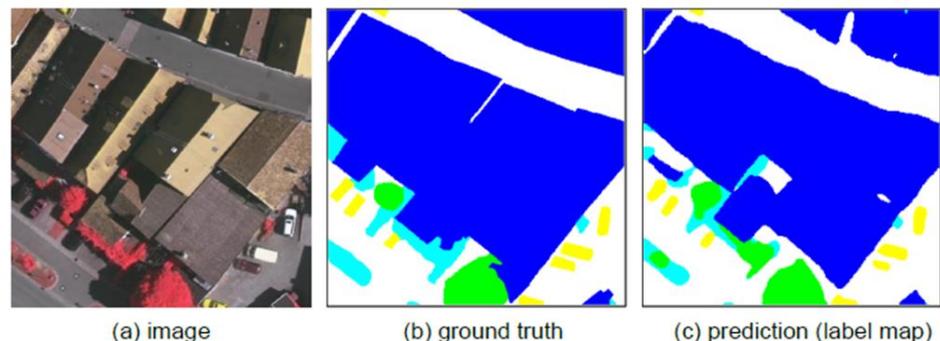


UNet



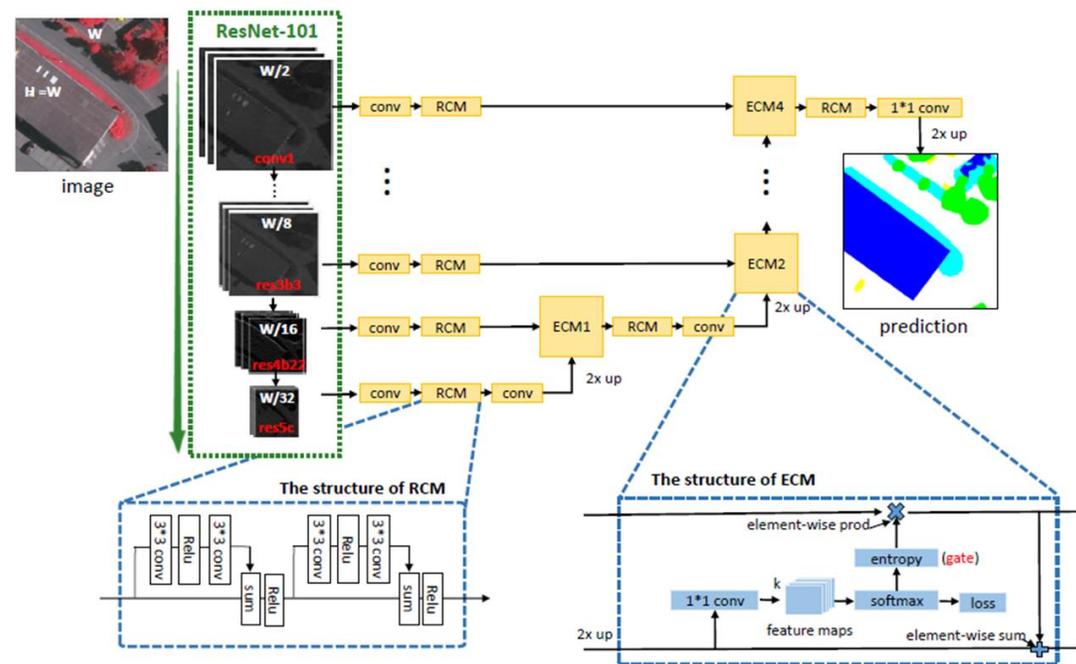
Skip connection系 [Wang et al., 2017]

- CNNの浅い層が持つLow-level featureの中には不要な情報(ノイズ)も多く含まれる
- 単純にスキップさせるのではなく、有用な情報を取捨選択する必要がある

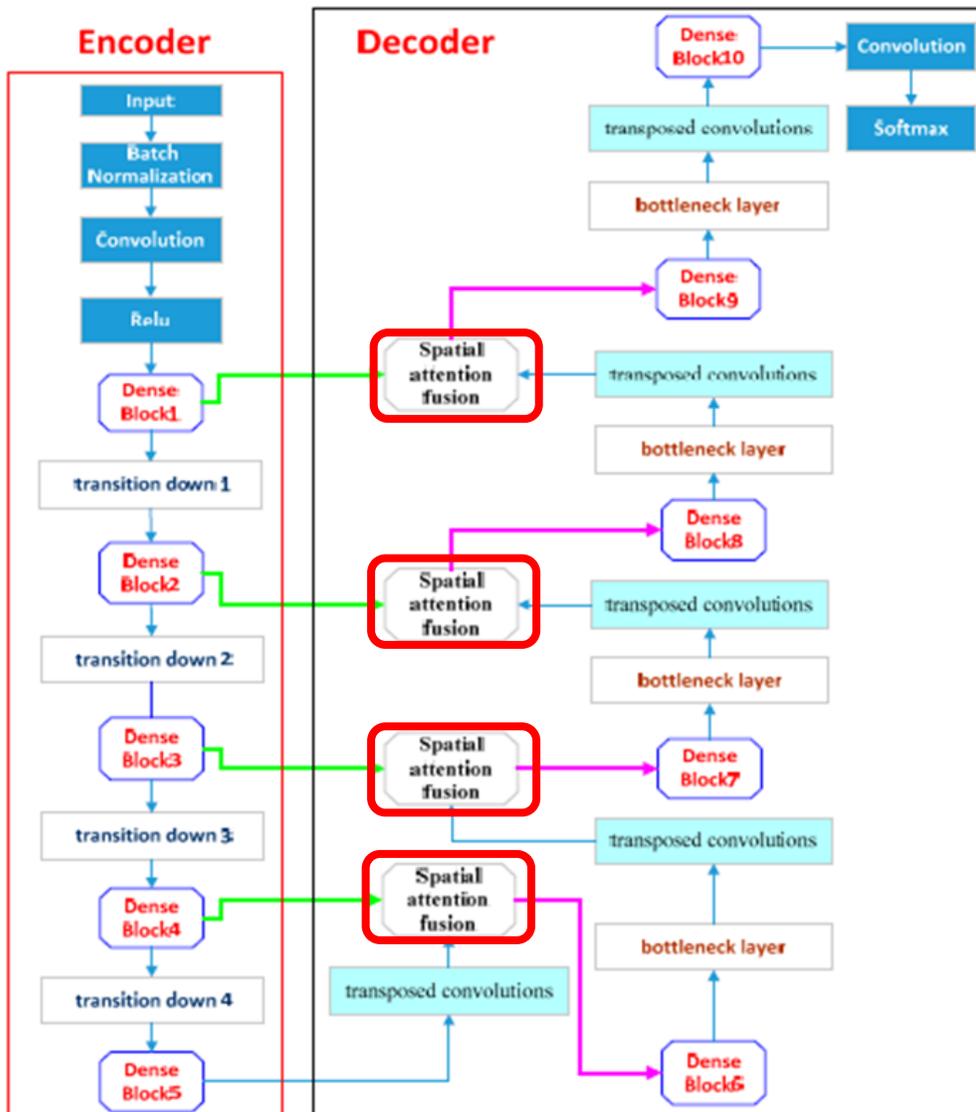


エラーマップ 予測のエントロピー

- CNNはオブジェクトの境界付近で自信がない(予測のエントロピーが高い)
- 各段階で予測結果を出力し、自信がないところに絞って低層から情報を持ってくる



Skip connection系 [Yang et al., 2018]



取捨選択をどうやるか？

→ Attentionマップを直接CNNで推定してあげる



(a) Original (dataset 2)



(b) Prediction (dataset 2)

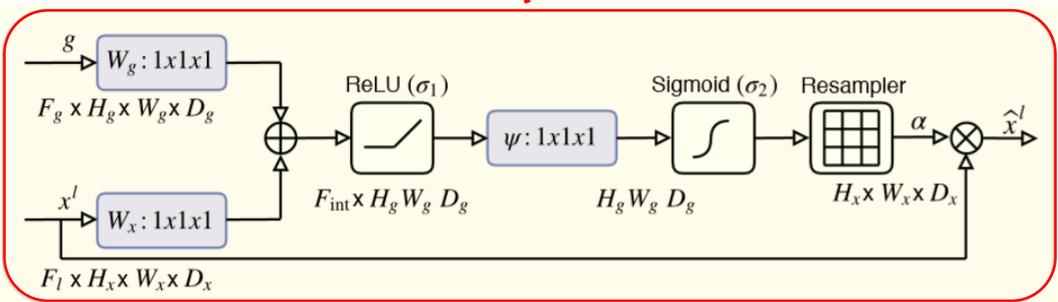
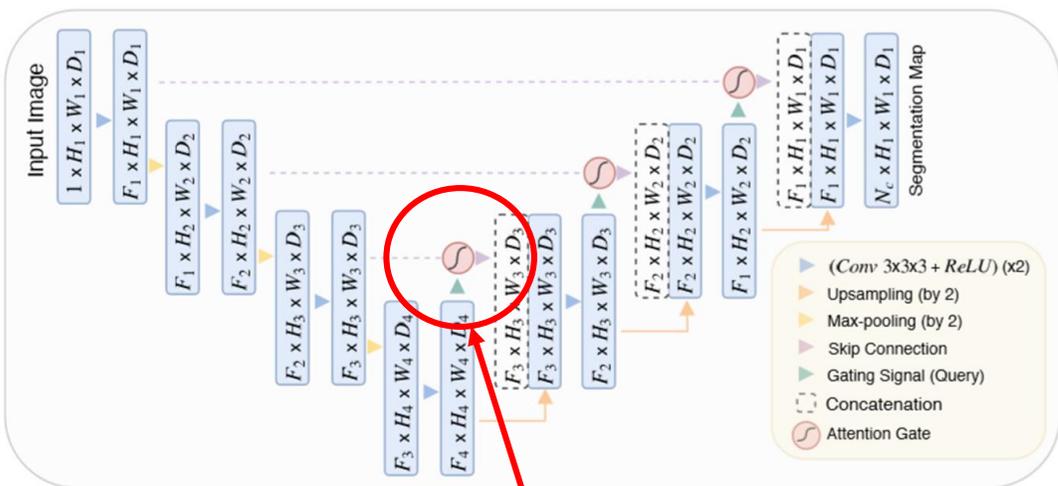


(c) Ground truth (dataset 2)

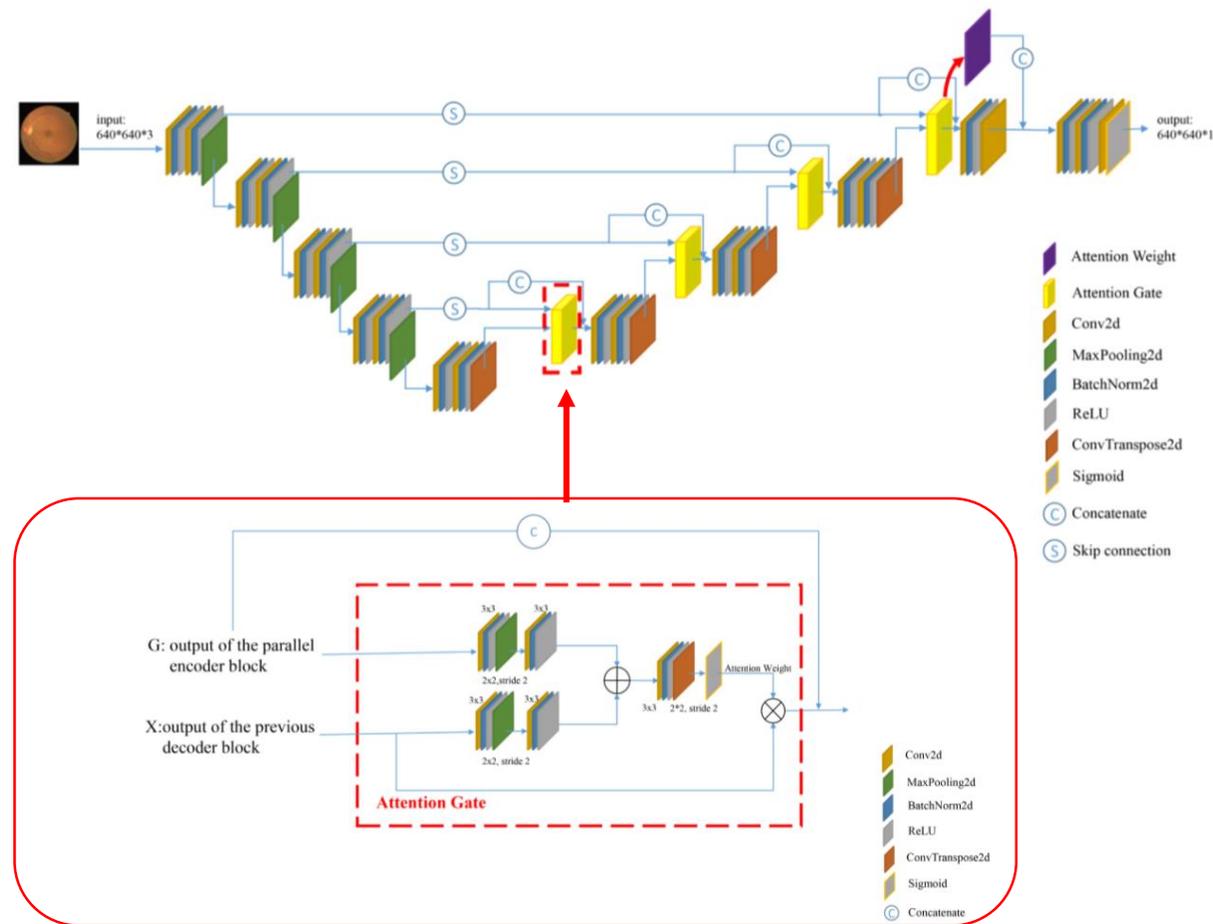
Attention UNet

本家本元のBiomedical系の分野でもSkipする特徴の取捨選択を行う手法が出てきている

(Oktay et al., 2018)



(Li et al., 2019)



まとめると

リモートセンシング画像の認識では・・・

- 解像度を落としたくない
- 受容野も広げたい



これをどう両立するかが課題



Skip connectionを用いる手法

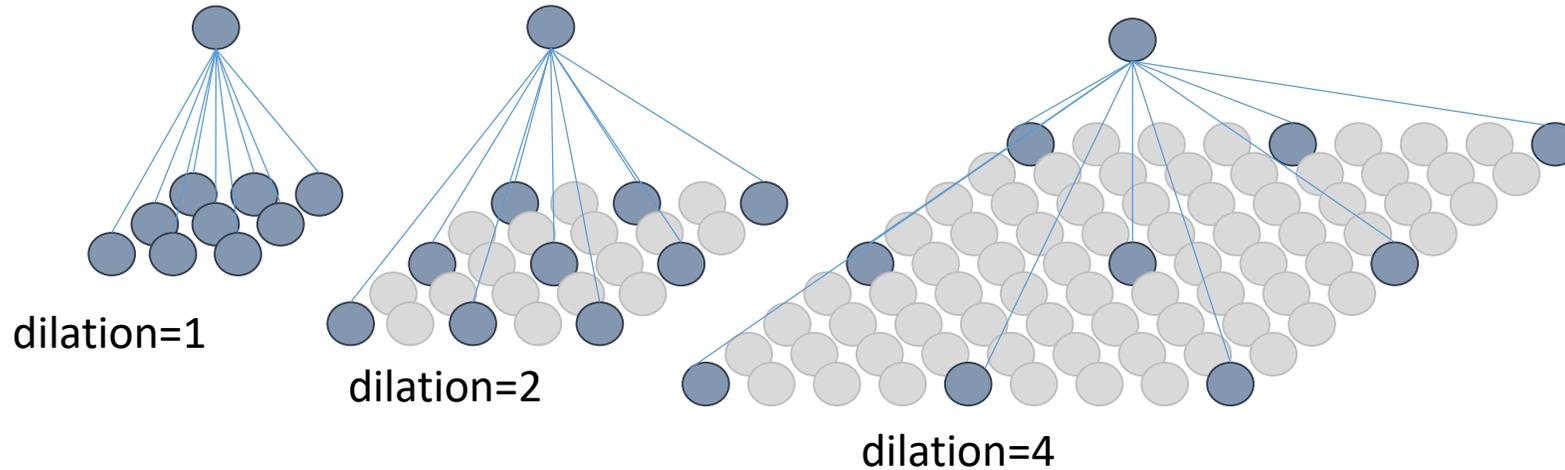
- ✓ 受容野はPoolingで広げる
- ✓ 浅い層の特徴を再利用することで失った解像度を回復



Dilated convolutionを用いる手法

- ✓ 解像度を落とさずに受容野を広げることのできる特殊な畳み込みカーネルを用いる

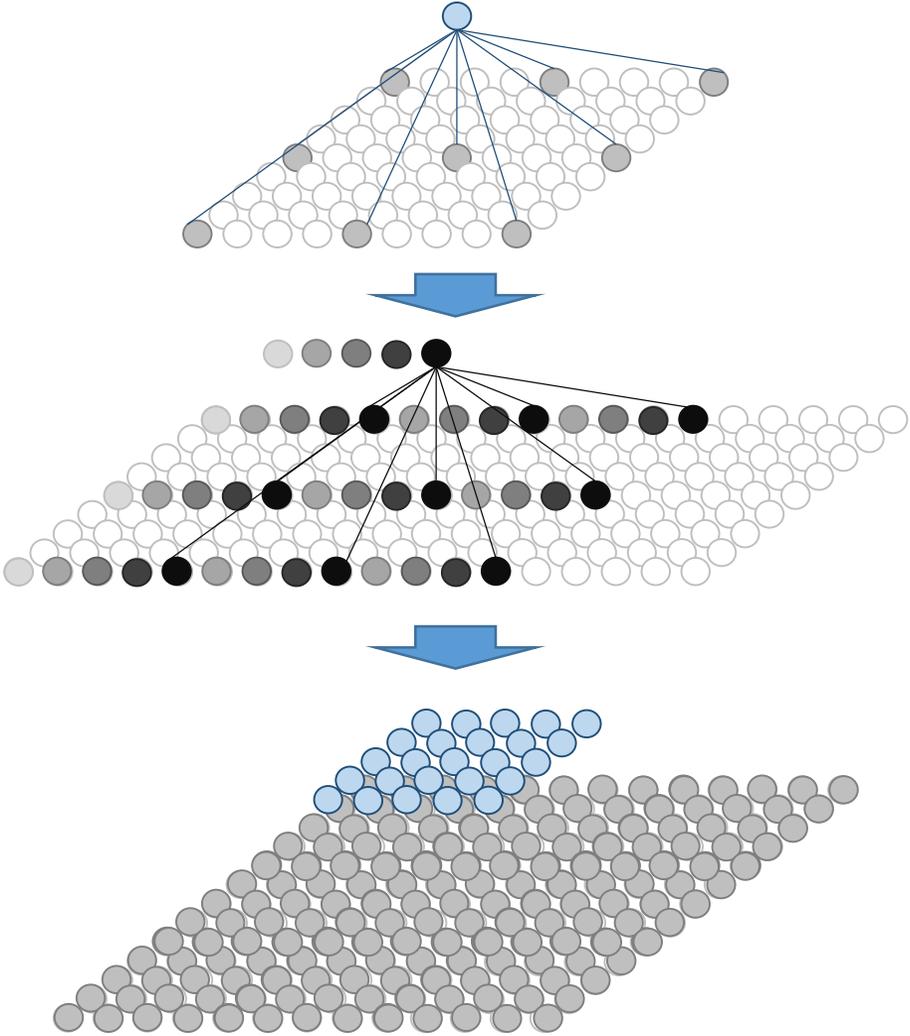
Dilated convolution



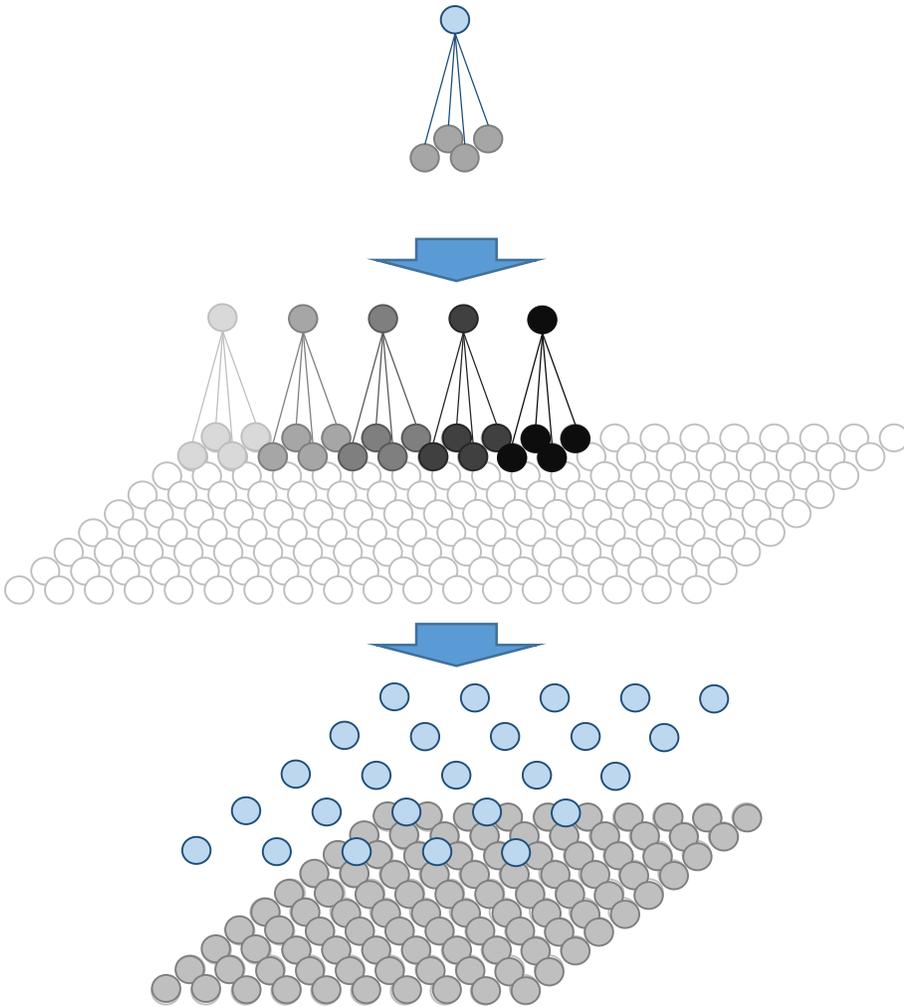
- 畳み込みカーネルの間隔を広げることで受容野を広げる方法
- 受容野を広げるのにサブサンプリングが必要ないので解像度が下がらない

Dilated convolution vs. Pooling

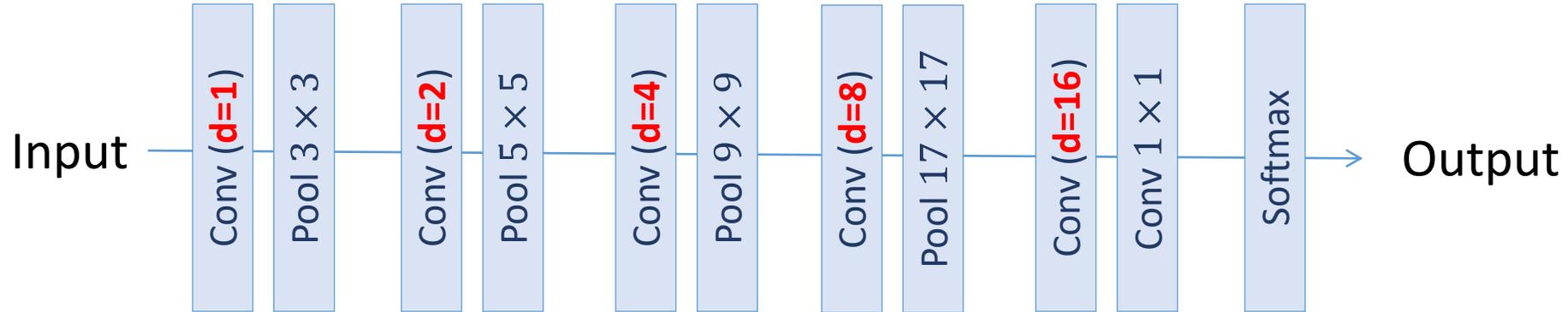
Dilated convolution



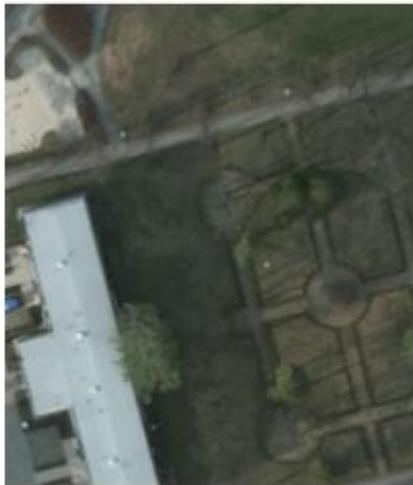
Pooling



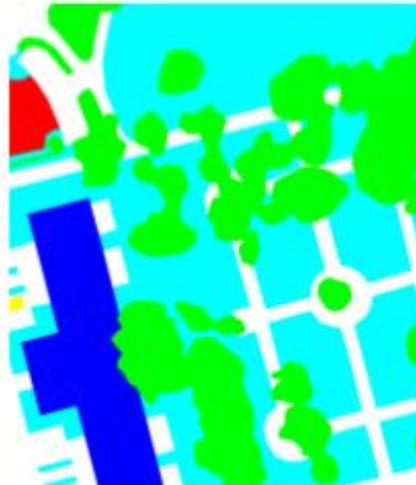
ダウンサンプリングの排除 [Sherrah, 2016]



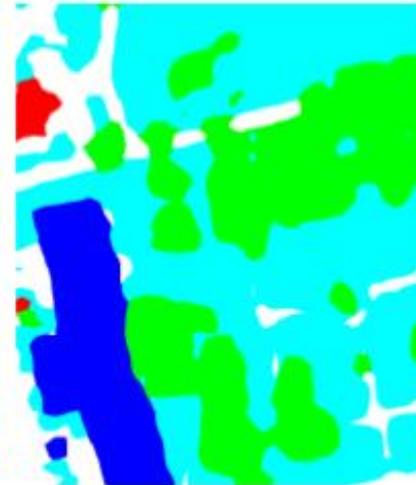
Pooling含めストライドはすべて1



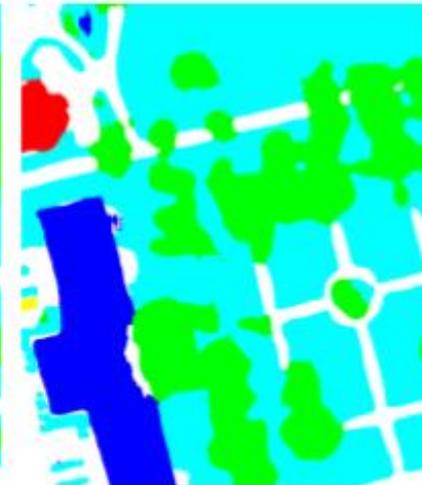
Input



GT

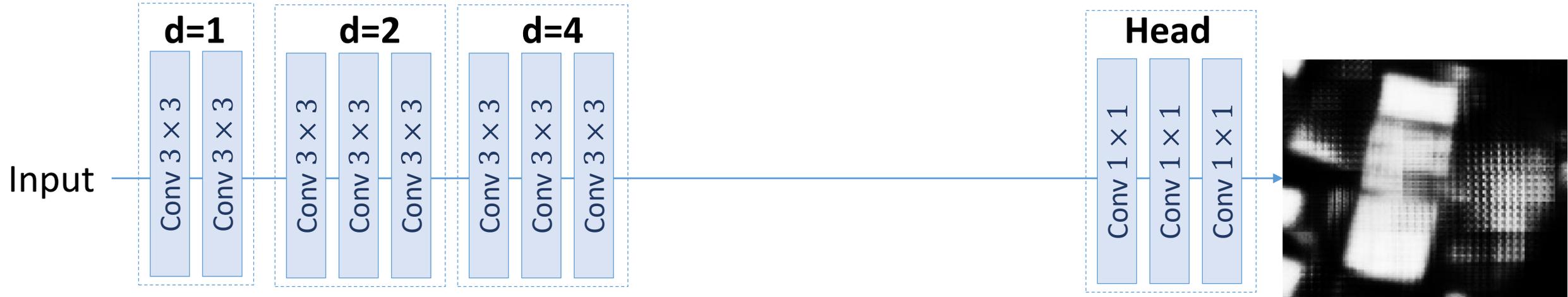


Downsampling-FCN



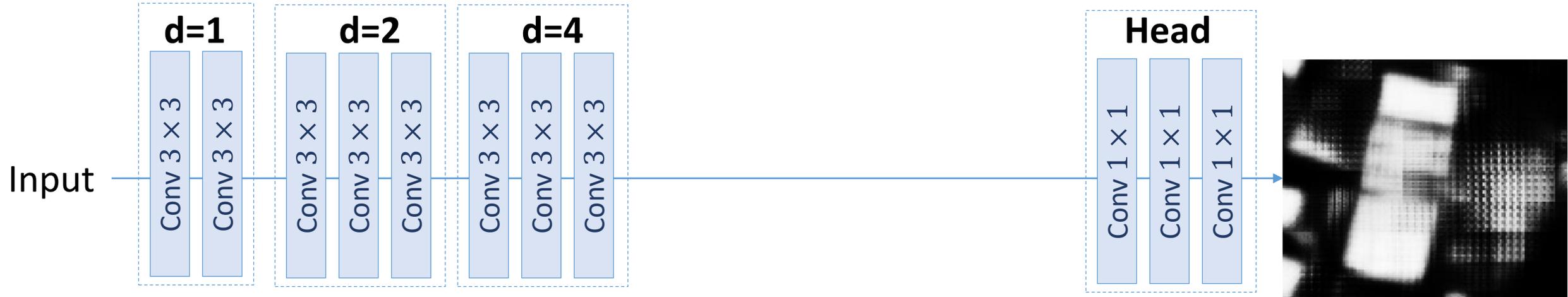
No-downsampling-FCN

ダウンサンプリングの排除 [Hamaguchi et al., 2018]

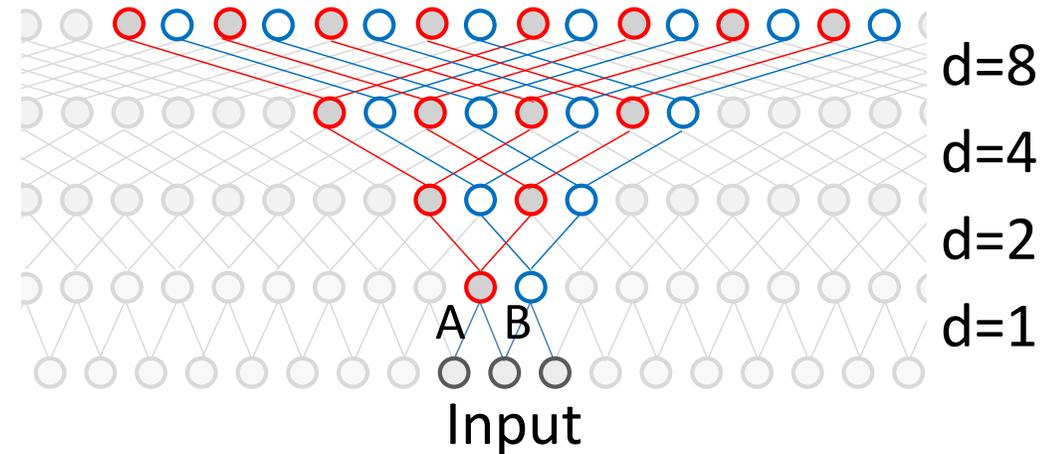
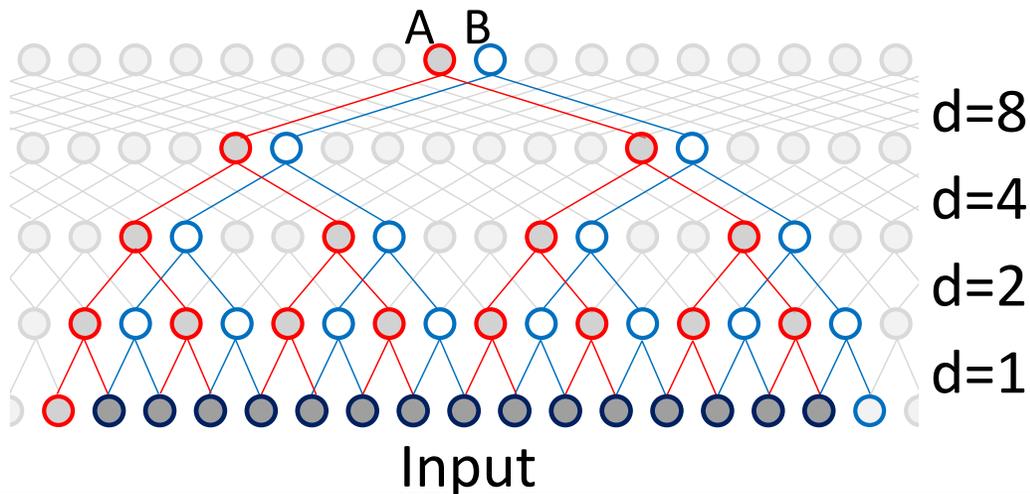


Dilationを増大させていくと出力にグリッド状のノイズが出る

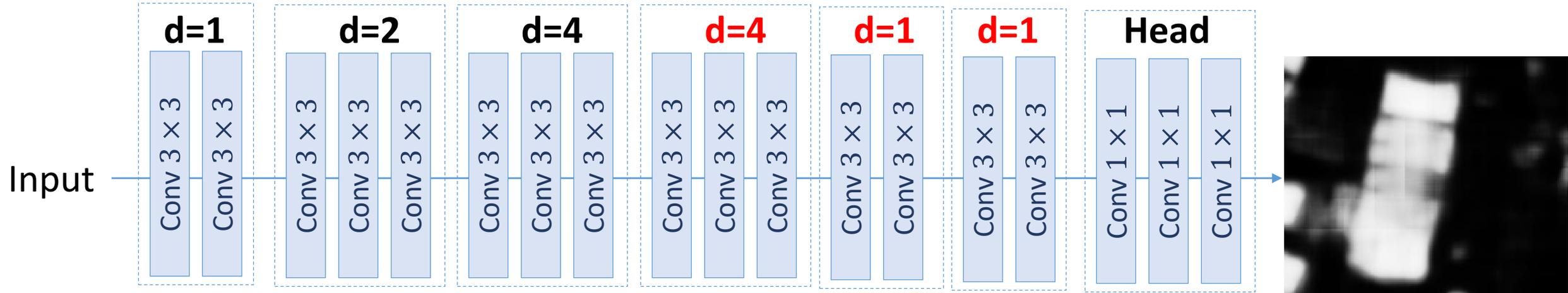
ダウンサンプリングの排除 [Hamaguchi et al., 2018]



ノイズの原因: 隣接するピクセル間で中間層の特徴を共有できていなかった

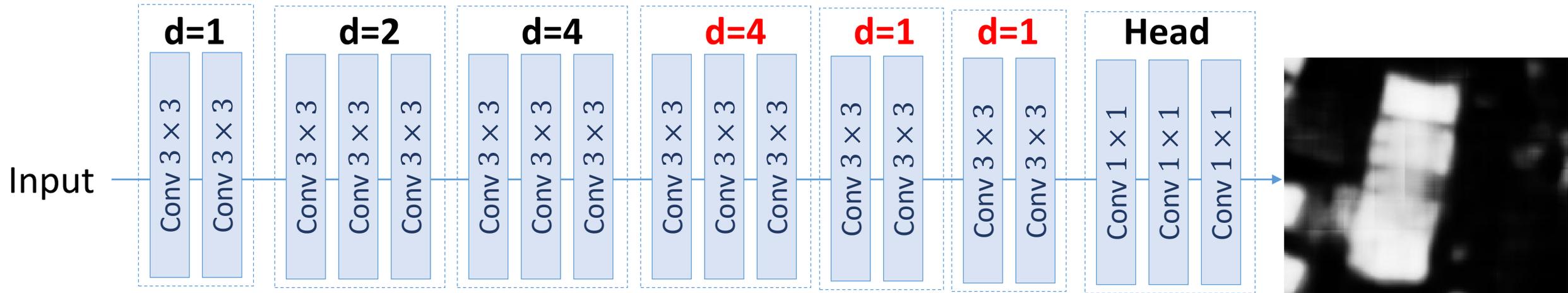


ダウンサンプリングの排除 [Hamaguchi et al., 2018]

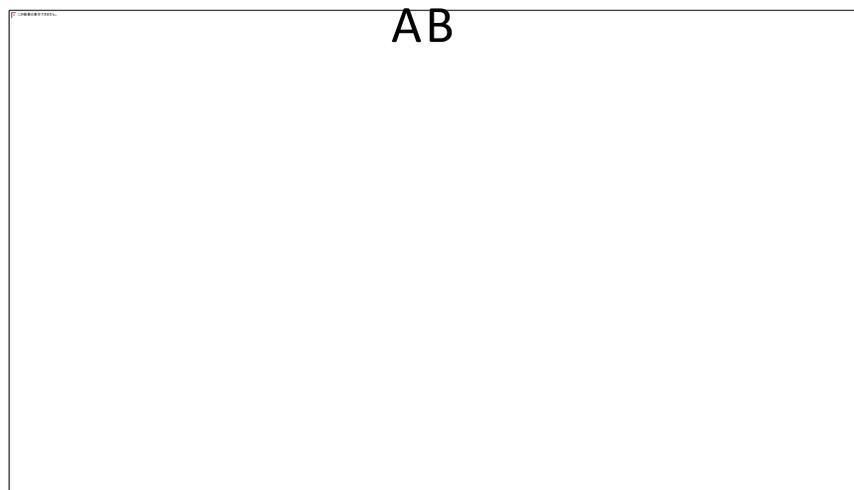


Dilationを増大させた後に減少させると出力のノイズが消える

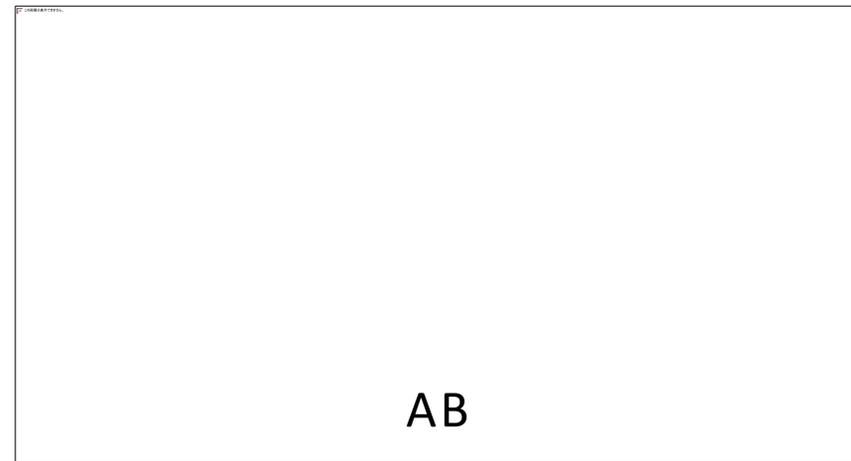
ダウンサンプリングの排除 [Hamaguchi et al., 2018]



中間層の情報がうまく共有されている



Input



ダウンサンプリングの排除 [Hamaguchi et al., 2018]



FCN-8s



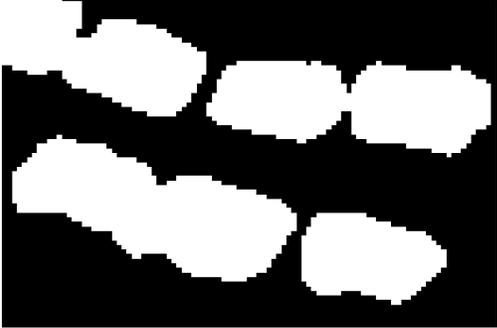
DeepLab-LFOV



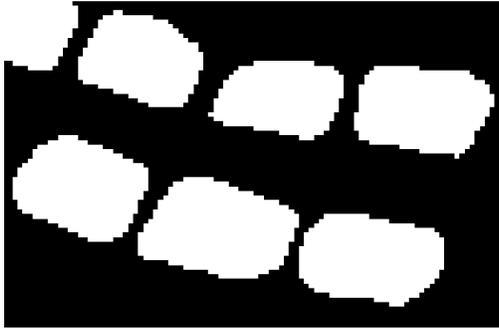
DeepLab-V2



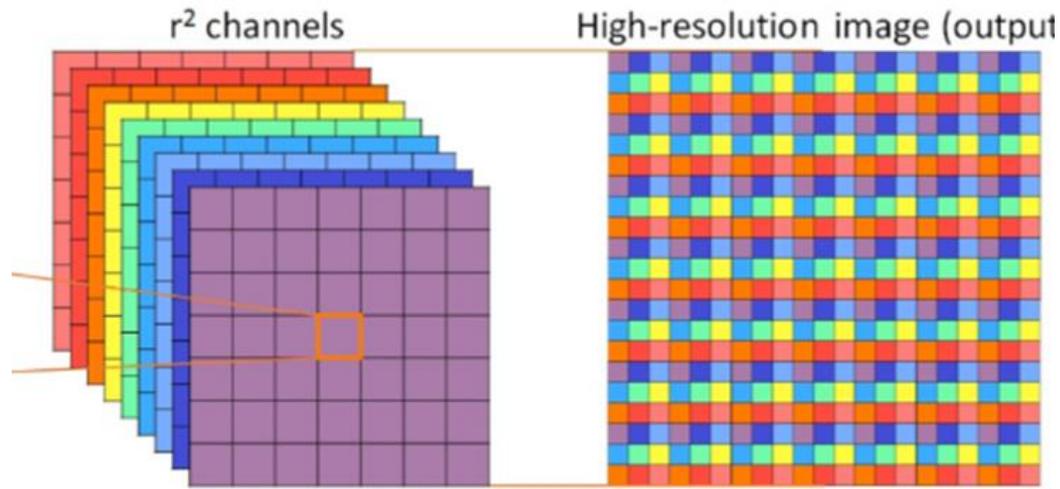
UNet



提案手法

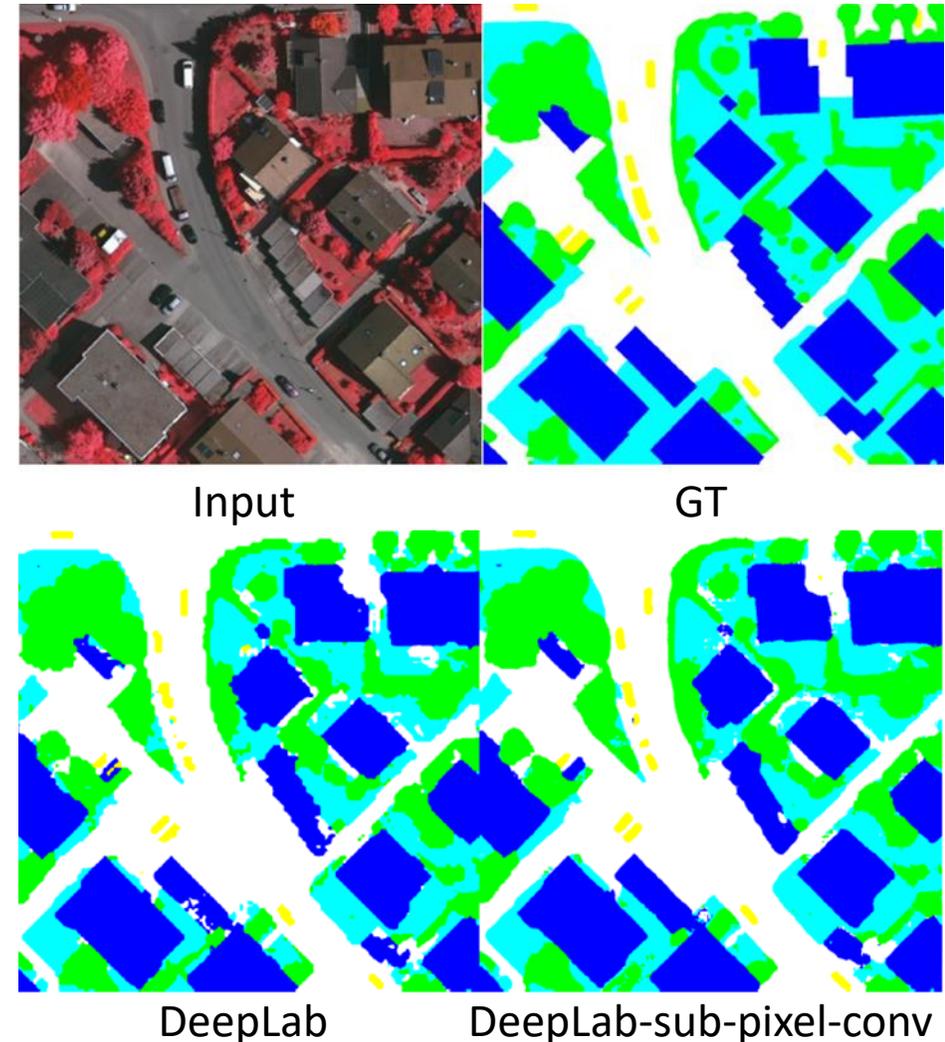


高解像度情報をCNNのチャネル方向で保持する [Chen et al., 2018]



Sub-Pixel Convolution [Shi et al., 2016]

- 複数のチャネルの特徴を順番に並べて高解像度のモザイクにUpsampleする方法
- 超解像の文脈で提案された手法
- リモートセンシング画像のセグメンテーションの問題に応用



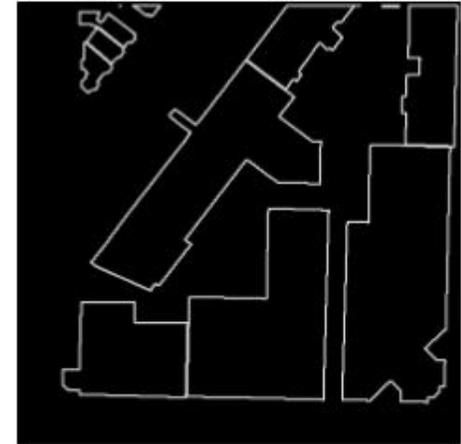
Chen, K., Fu, K., Yan, M., Gao, X., Sun, X., & Wei, X. (2018). Semantic Segmentation of Aerial Images With Shuffling Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 173–177.

Boundary-awareなLoss関数 [Yuan, 2016]

- 建物か建物でないかのバイナリではなく、建物の境界までの距離を推定する問題を設定(d)
- 境界までの距離をGTとすることでオブジェクトの境界や形状に関してより豊富な情報をCNNに与えることができる



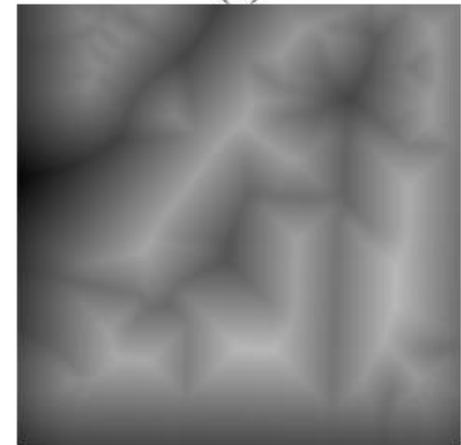
(a)



(b)



(c)

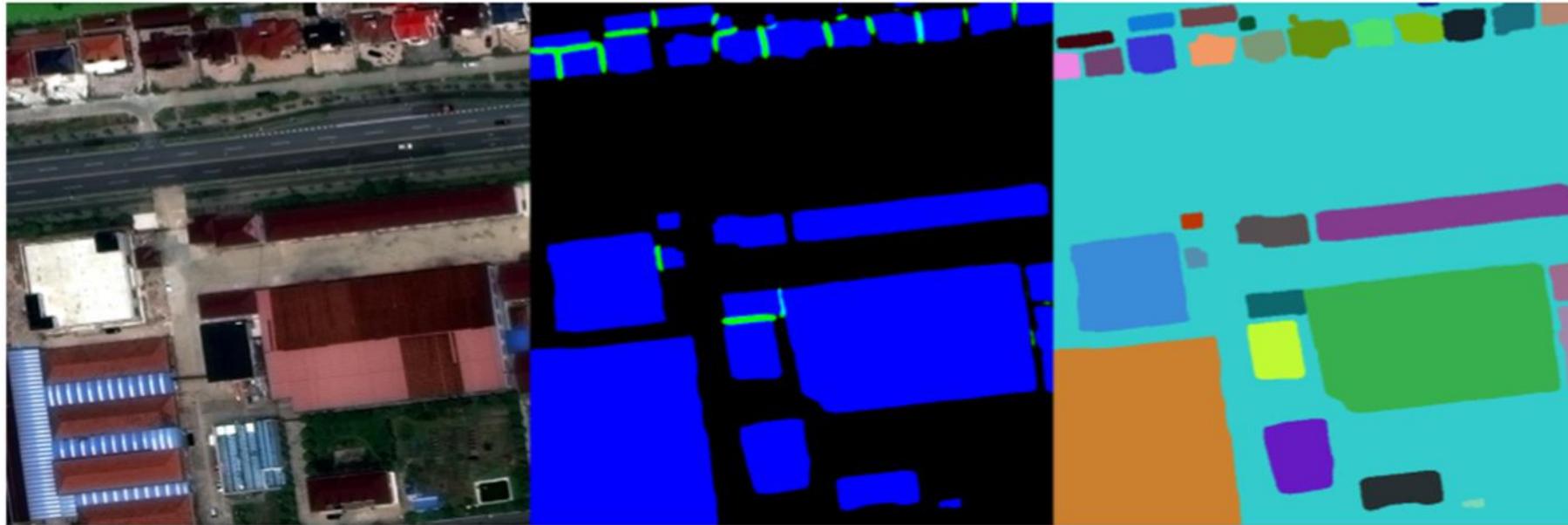


(d)

Yuan, J. (2016). Automatic Building Extraction in Aerial Scenes Using Convolutional Networks. *ArXiv Preprint ArXiv:1602.06564*.

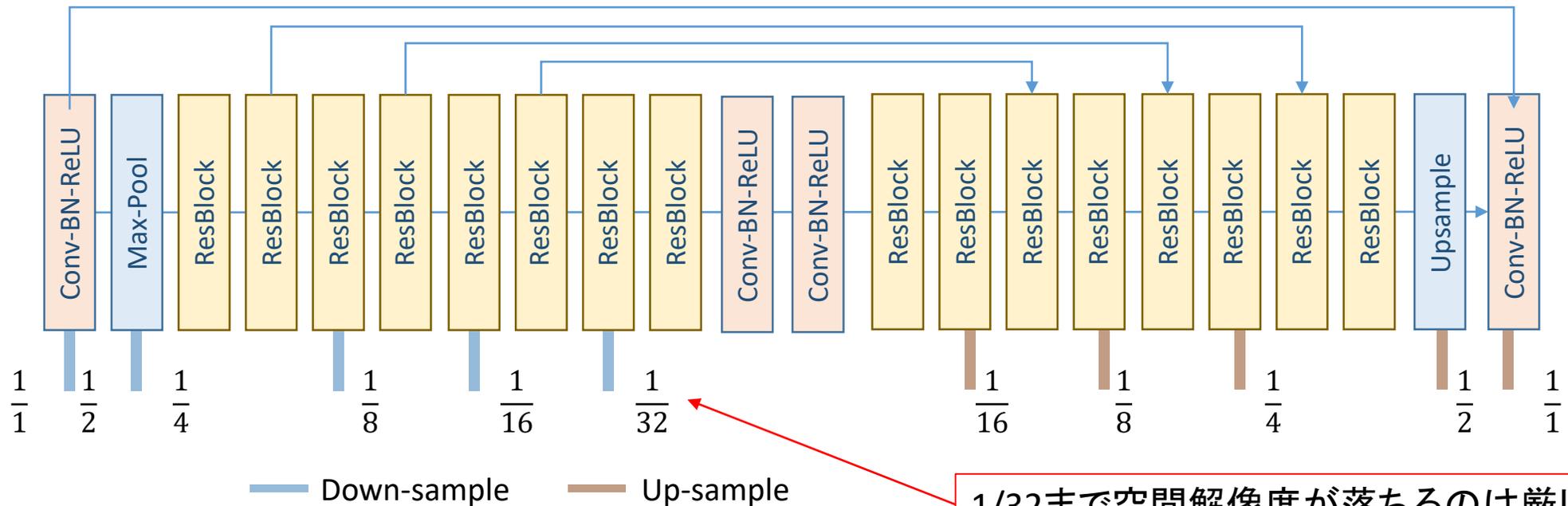
近接するオブジェクトの境界を明示的に推定 [Iglovikov et al., 2018]

- Skip connectionを用いても近接する建物がくっついて検出されてしまう
- 建物を1棟ずつ分離して検出するため、ネットワークに近接する境界を検出するブランチを追加
- 検出された境界を用いて建物検出結果を後処理



TIPS : 特徴マップの解像度に注意を払う

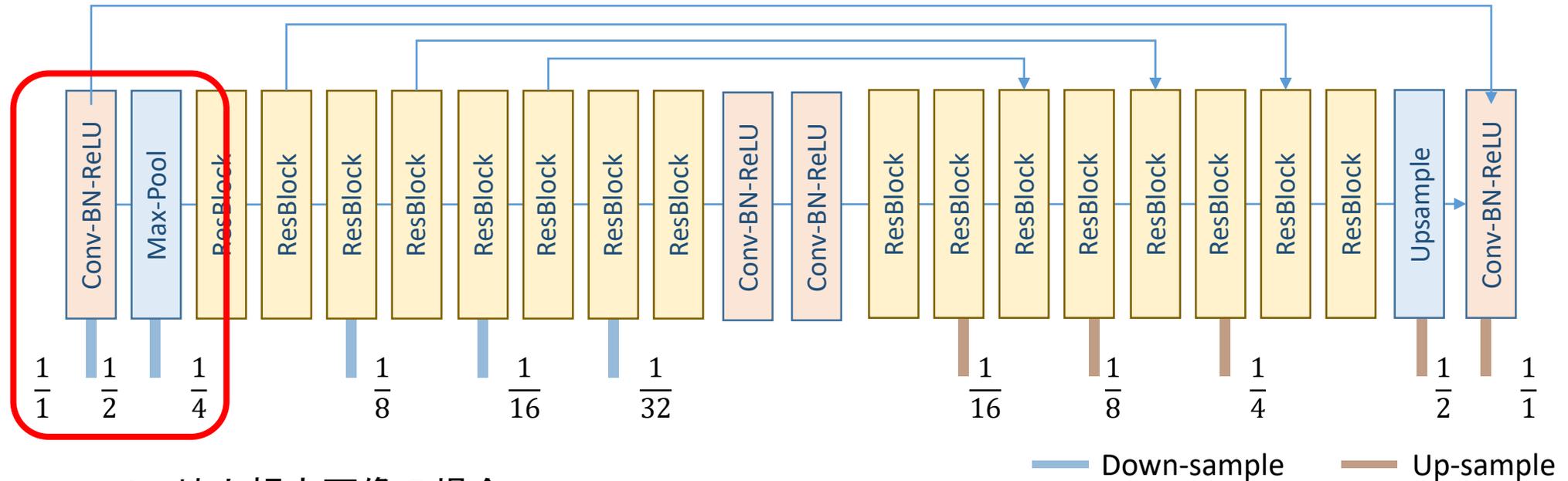
解像度に注意を払ってネットワーク設計をするだけで、お手軽に精度が上がる場合がある



ResNet18をバックボーンにしたResidual UNet

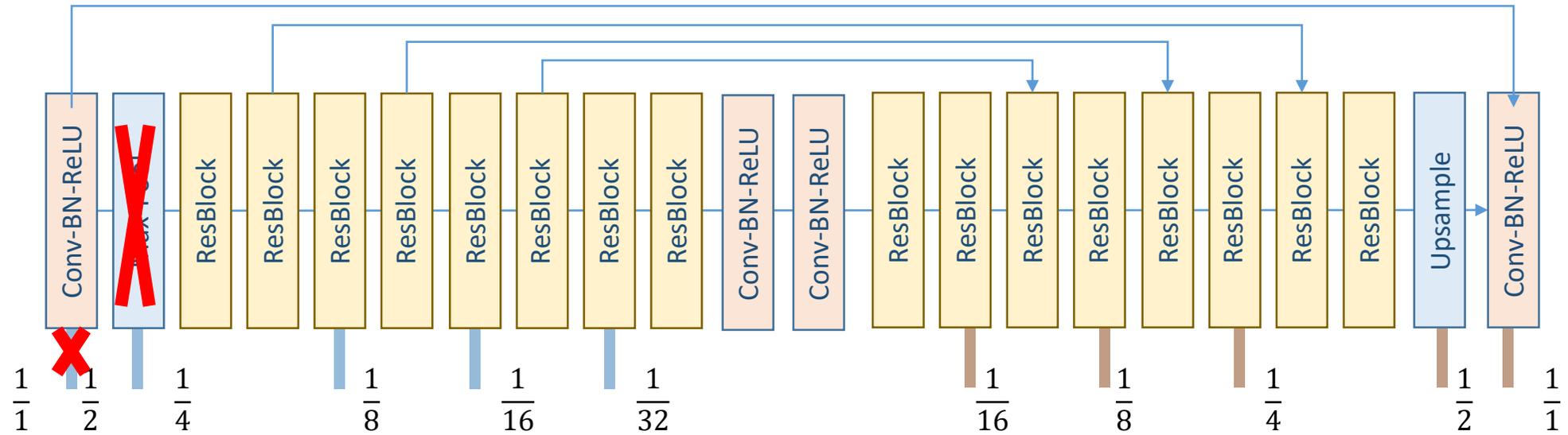
TIPS : 特徴マップの解像度に注意を払う

特にネットワークの入り口で1/4に落ちるのは情報のロスが大きい



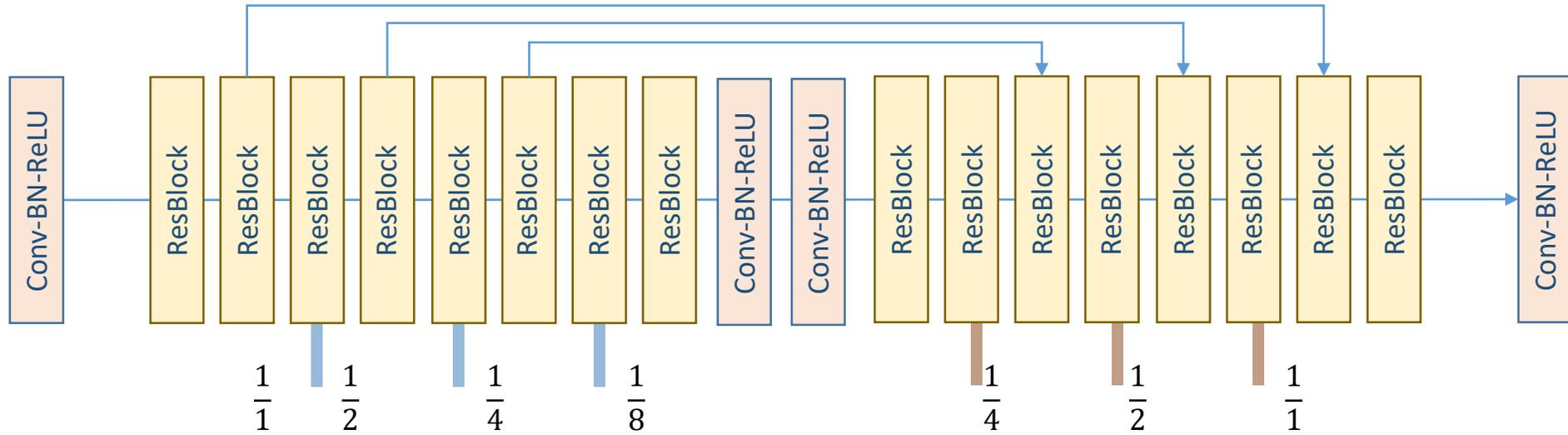
- 地上視点画像の場合・・・
 - 個々のオブジェクトを認識するのに十分な解像度がある
 - ネットワークの入り口で積極的にダウンサンプルしてから深くする
 - メモリ効率が良いので理にかなっている
- リモートセンシングの場合・・・
 - 対象とするオブジェクトが小さいので解像度が命
 - ダウンサンプルする前に十分に情報処理しておく必要がある

TIPS : 特徴マップの解像度に注意を払う

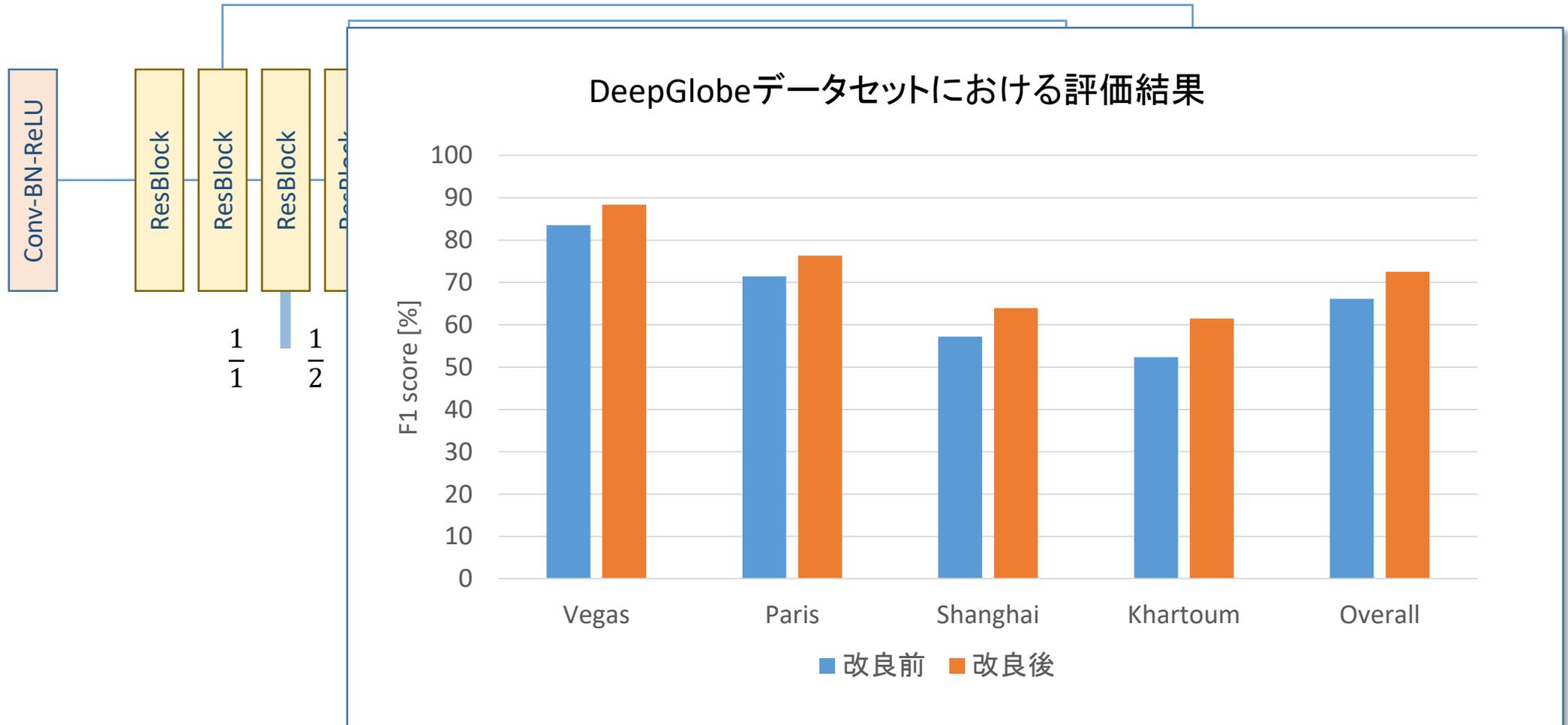


入口付近のダウンサンプリングを2つ削除

TIPS : 特徴マップの解像度に注意を払う



TIPS : 特徴マップの解像度に注意を払う



本日の内容

- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - **道路検出はトポロジーまで考える**
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

道路検出タスクにおける特有の課題

街路樹等によるオクルージョンによって検出結果が途切れてしまう



入力画像ピクセルの分類結果としては小さな間違い
道路の接続関係が変わってしまうので、地図としては大きな間違い



課題: 道路をトポロジーを含めて正しく検出するにはどうすればいいか?



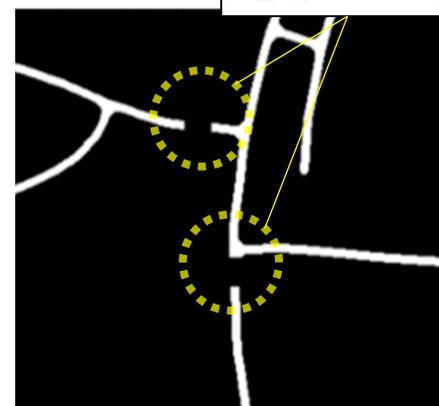
影や街路樹によるオクルージョン [Bastani et al., 2018]



Input



GT

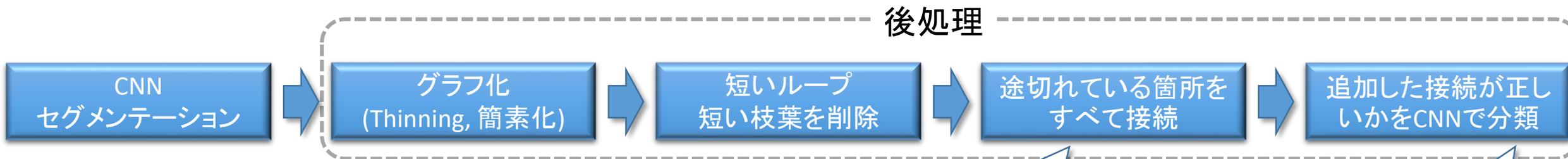


推定結果

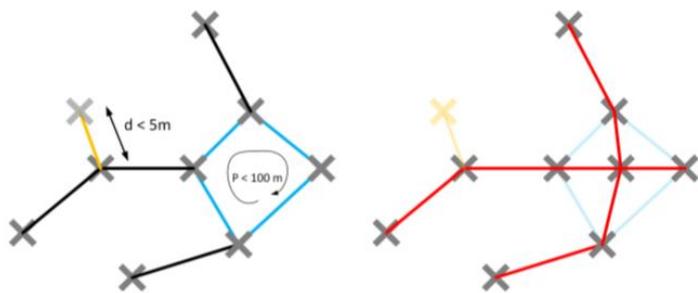
途切れている

[Hu et al., 2019]

後処理による対処 [Máttyus et al., 2017]



- 5m未満のエッジを持つ端点は削除
- 100m未満のループは削除



- 全ての端点(行き止まり)について処理
- 半径50m以内にある他のノードと接続
- 接続先が複数ある場合は距離などを基に決定
- 接続可能な端点はすべて接続する



- 追加した接続の周囲の画像をCNNに入力
- 正しい接続であるか否かを識別させる

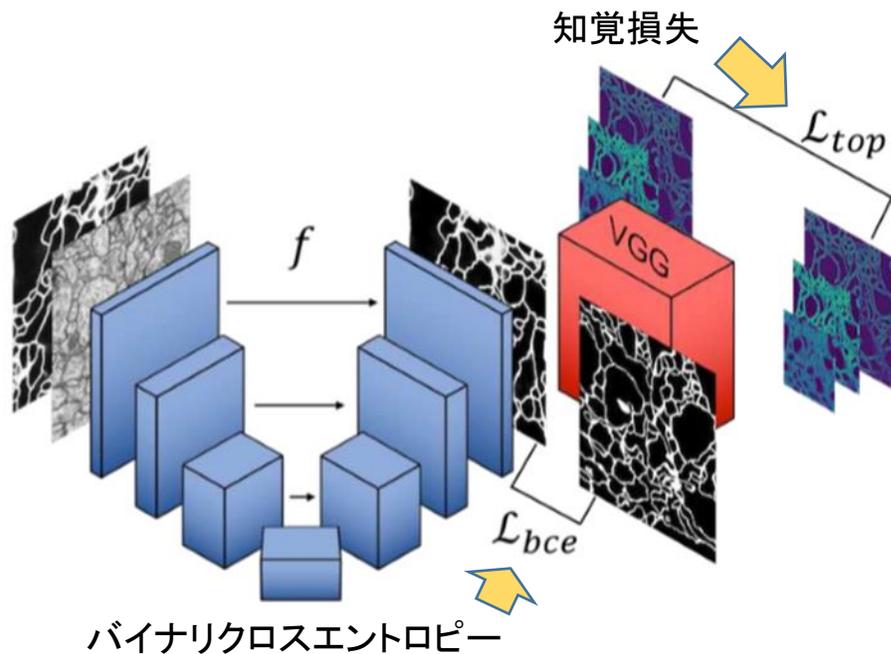


間違った接続

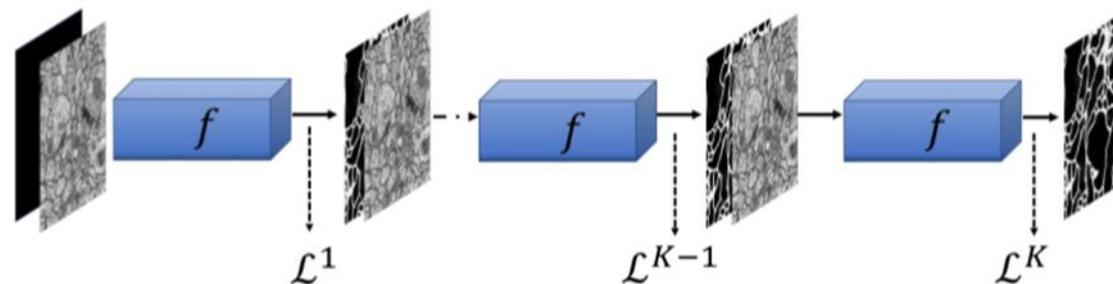
正しい接続

知覚損失の導入 [Mosinska et al., 2018]

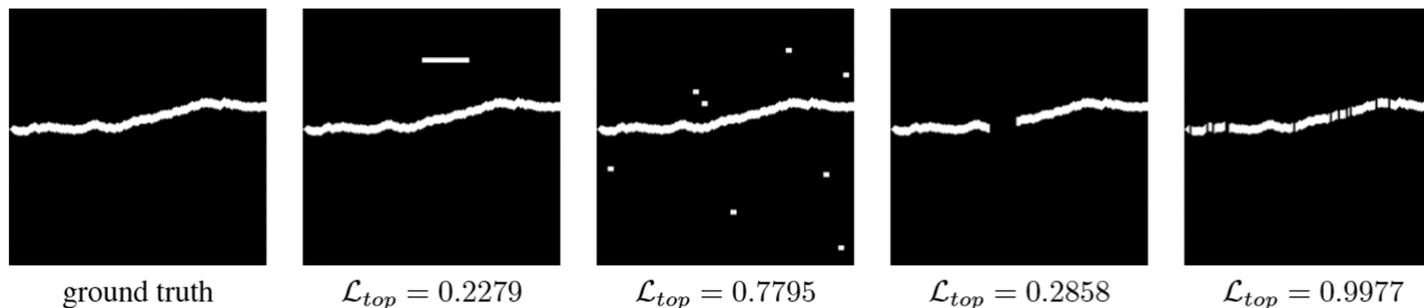
知覚損失を導入することでトポロジーの違いに感度の高い学習を行う



繰り返し処理による道路マップの精緻化



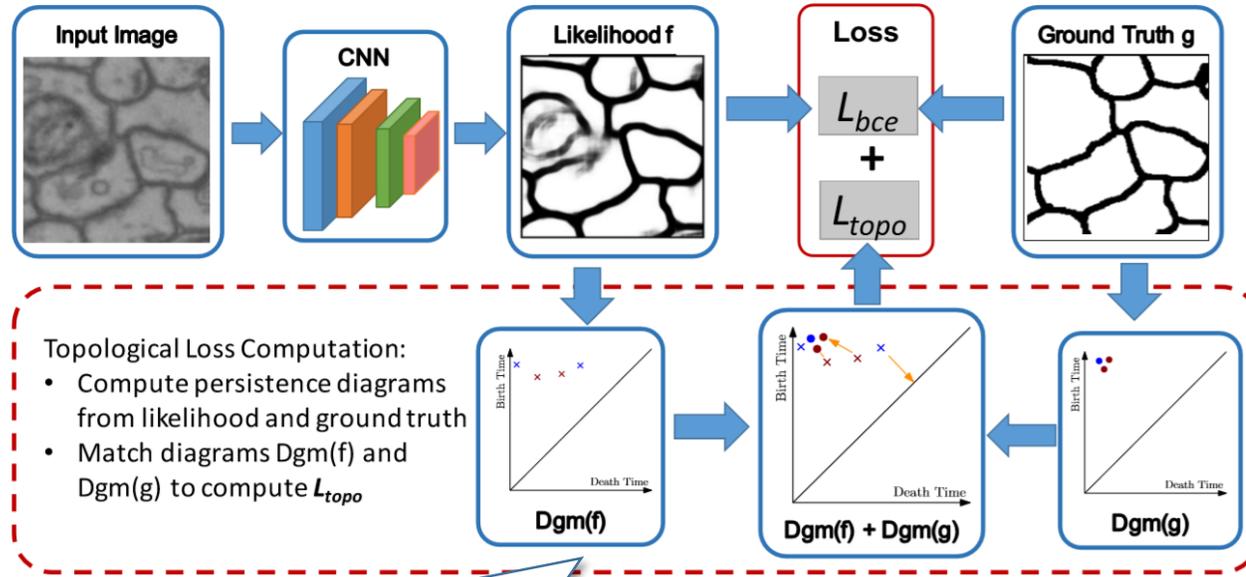
トポロジーの違いと知覚損失



Mosinska, A., Marquez-Neila, P., Kozinski, M., & Fua, P. (2018). Beyond the Pixel-Wise Loss for Topology-Aware Delineation. CVPR.

計算トポロジーの応用 [Hu et al., 2019]

計算トポロジーの手法を用いることで、Ground Truthとのトポロジーの違いをより直接的に損失関数に組み込んだ

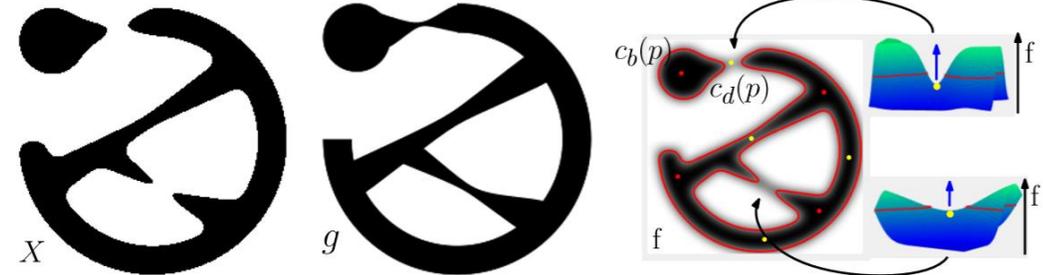


Persistence diagram

二値化閾値を1 → 0に動かしたときのトポロジー構造の生成消滅の履歴
GTのDiagramとの差分を損失とすると、トポロジーの変化が起こるピクセルをGTに近づけるような損失関数が得られる

ざっくり何をしているか...

計算トポロジーを用いれば正解とのトポロジーの差を引き起こしているピクセルを特定できる原因となっているピクセルだけをGTに近づけるように学習している

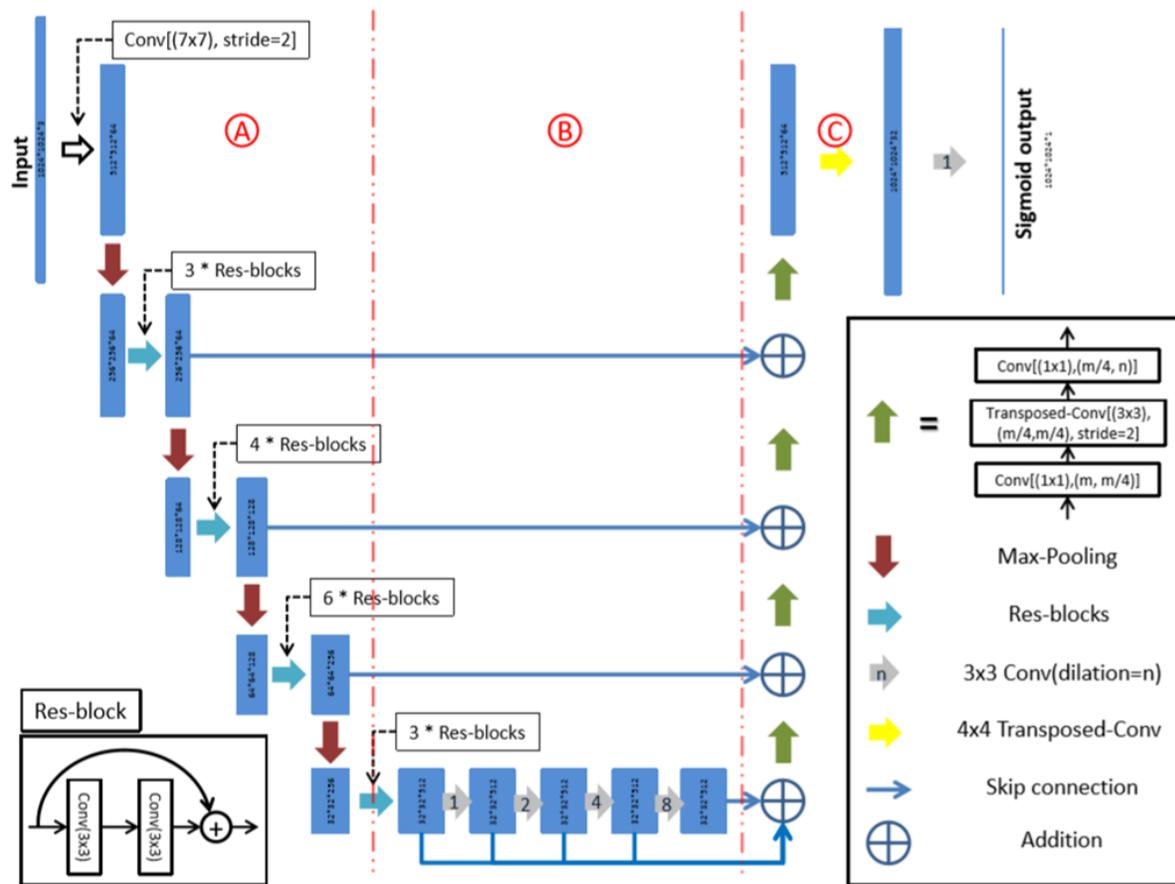


Hu, X., Fuxin, L., Samaras, D., & Chen, C. (2019). Topology-Preserving Deep Image Segmentation. *CoRR*.

受容野を広げる [Zhou et al., 2018]

DeepGlobe@CVPR2018における道路検出部門の優勝モデル

中央部分にDilated convolutionを用いることでネットワークの受容野を広げ、接続性の問題を軽減



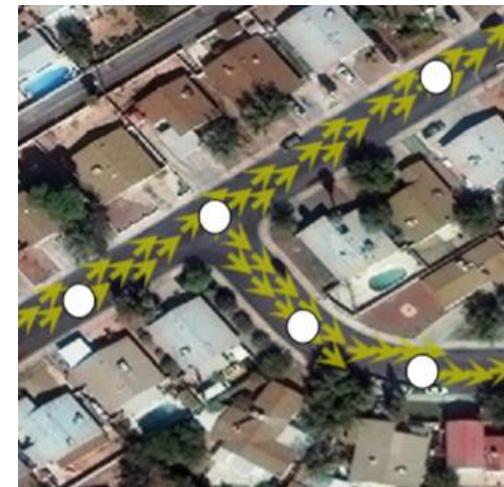
Zhou, L., Zhang, C., & Wu, M. (2018). D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. *CVPRW*.

道路の向きを学習する [Batra et al., 2019]

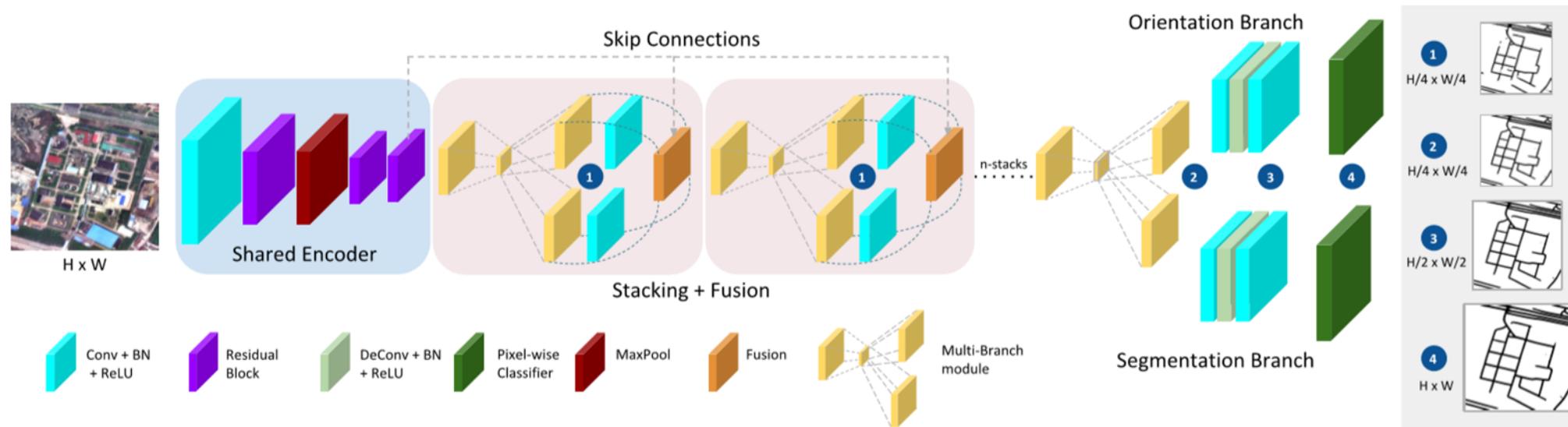
道路のセグメンテーションと同時に道路方向を推定するタスクを学習

向きを推定するには周囲の状況を含めて認識する必要がある

→ 向きを明示的に学習することでコンテキストを考慮した特徴量を効果的に学習している
SpaceNet, DeepGlobeの2つの大規模データセットで最高精度を達成している



セグメンテーションと向きの情報が混ざり合う

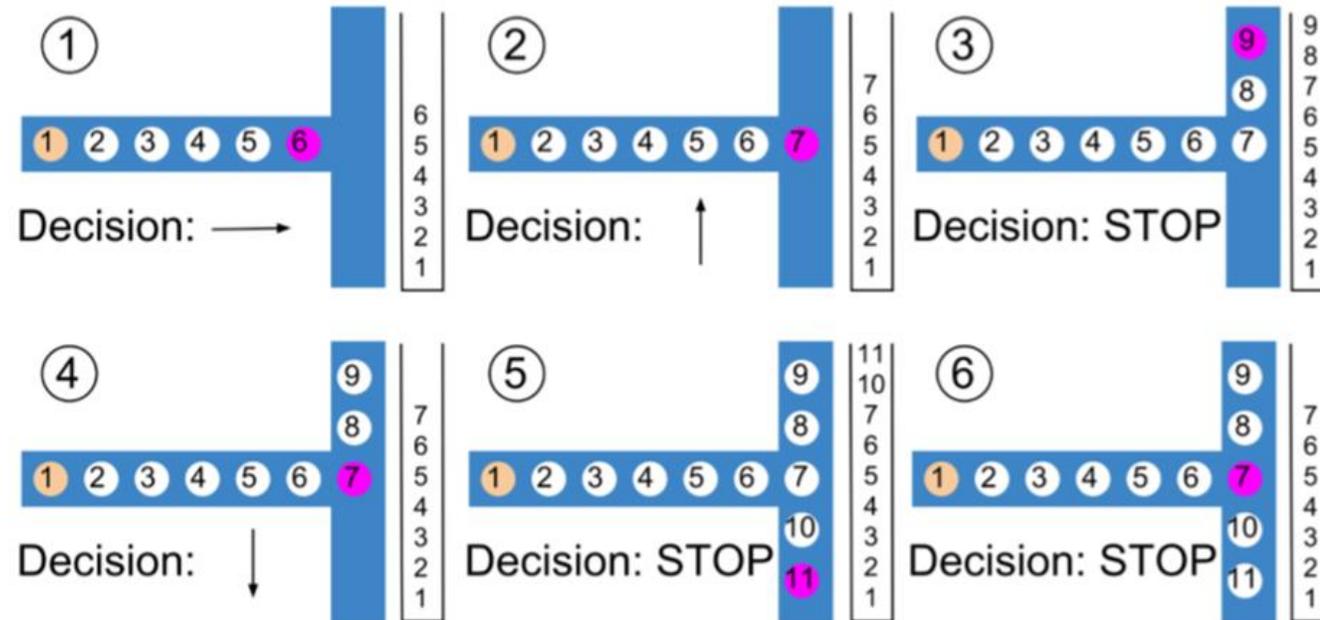


Batra, A., Singh, S., Pang, G., Basu, S., & Manohar, C. V. J. (2019). Improved Road Connectivity by Joint Learning of Orientation and Segmentation. *CVPR*.

セグメンテーションに依らない手法 [Bastani et al., 2018]

画像と生成済みマップを入力として、動的に道路マップを構築していくCNNを提案

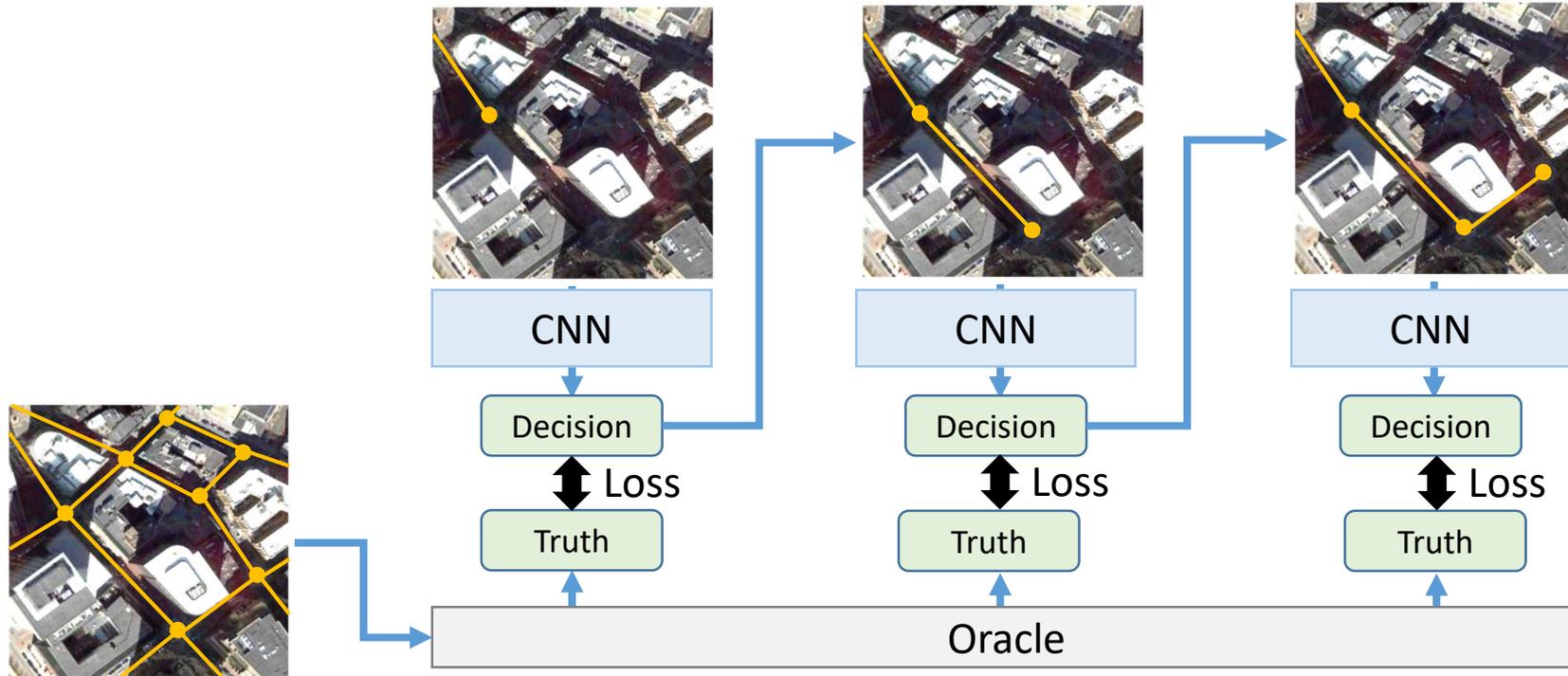
各ステップでCNNは停止または進行の行動と進行方向を出力する



進行: 一定距離進行してノードを生成し、エッジを張る
停止: 停止して前のノードに戻る

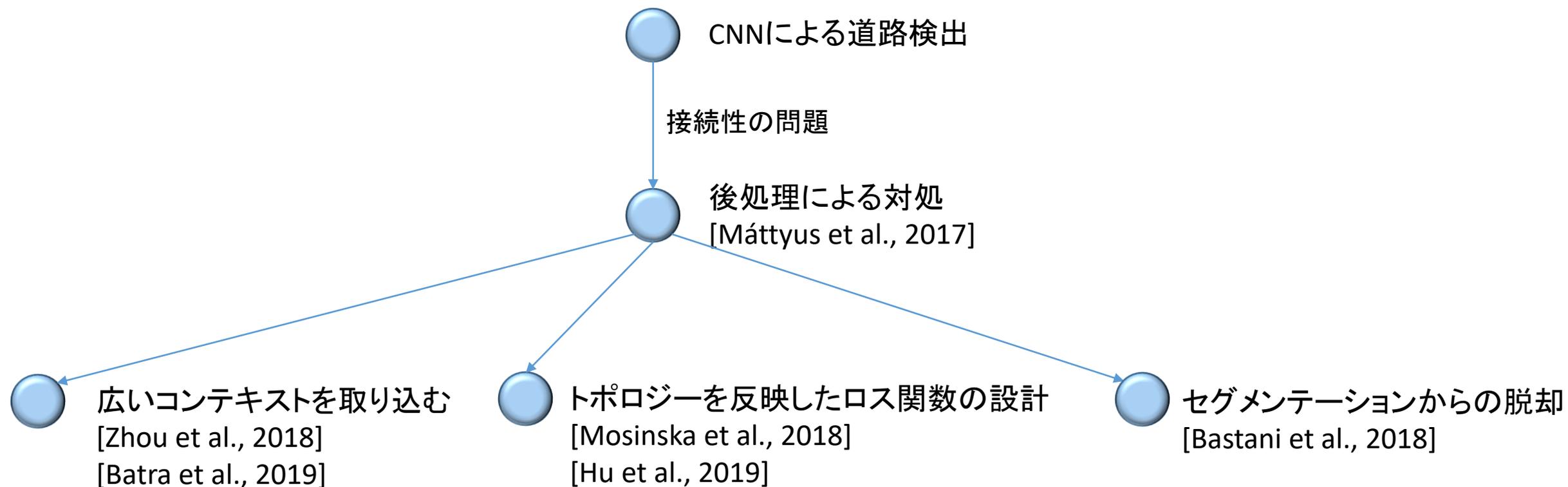
セグメンテーションに依らない手法 [Bastani et al., 2018]

オラクルは正解のマップを知っていて、各ステップでとるべき正解の行動を教えてくれる



Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S., ... Dewitt, D. (2018). RoadTracer: Automatic Extraction of Road Networks from Aerial Images. *CVPR*.

道路検出まとめ



本日の内容

- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - 道路検出はトポロジーまで考える
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

ご清聴ありがとうございました

株式会社パスコ



本日の内容

- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - 道路検出はトポロジーまで考える
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

ラベル付きデータの不足

- リモートセンシング画像に対する大規模なデータセットは不足している
- 特にマルチスペクトル画像に対しては使えるデータセットは限られてくる

- 少ないデータで学習できる手法が有用
 - データの豊富なドメインから知識を転移
 - 合成画像を用いた学習
 - 教師なし学習

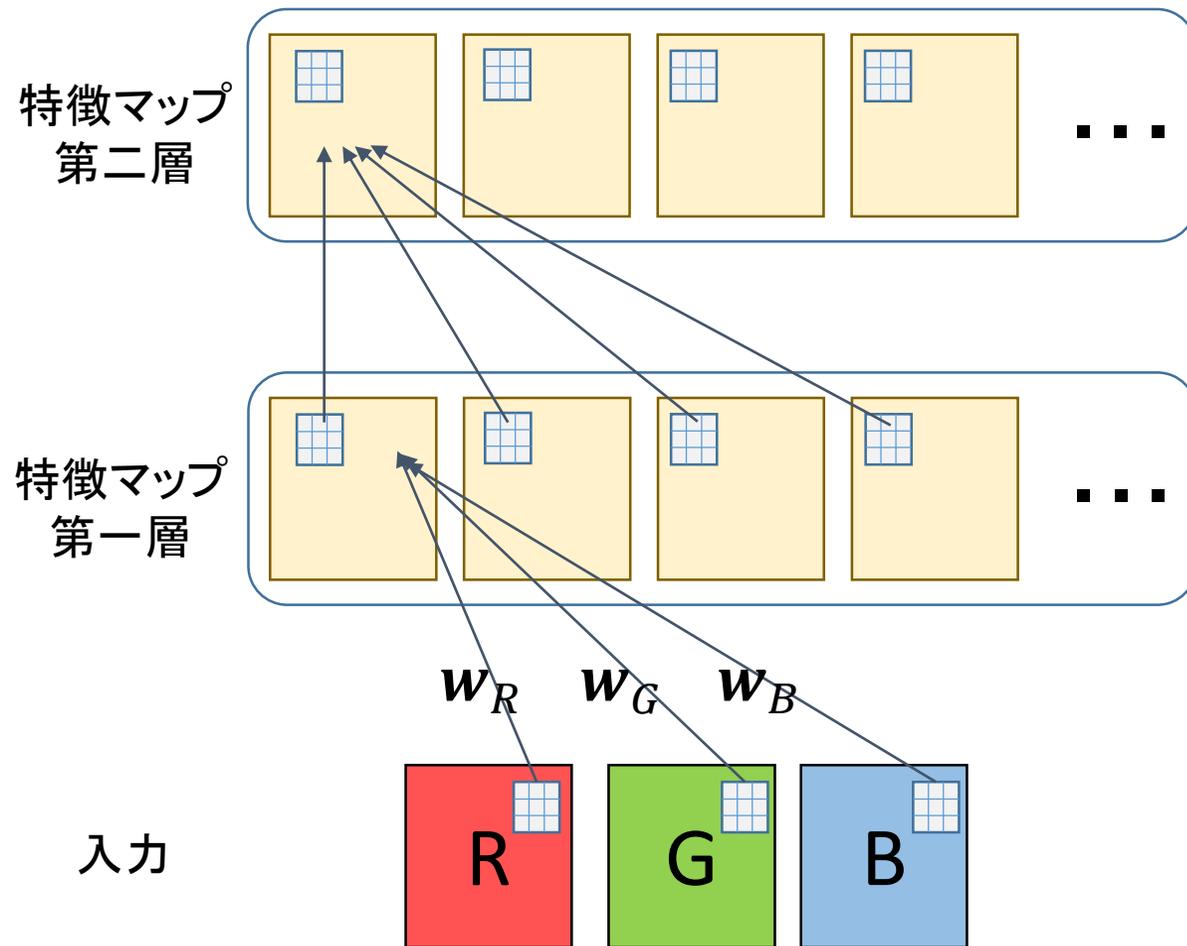


転移学習

- リモートセンシング画像にImageNet特徴は有用か？
 - 有効であることが報告されている [Penatti et al., 2015] [Marmanis et al., 2016]
- マルチスペクトル画像にどう対応する？
 - RSでは様々なスペクトルのデータを扱うので、単純にImageNetから転移できない場合も多い
 - RGBしか使わない or PCAで次元圧縮 → 情報のロス

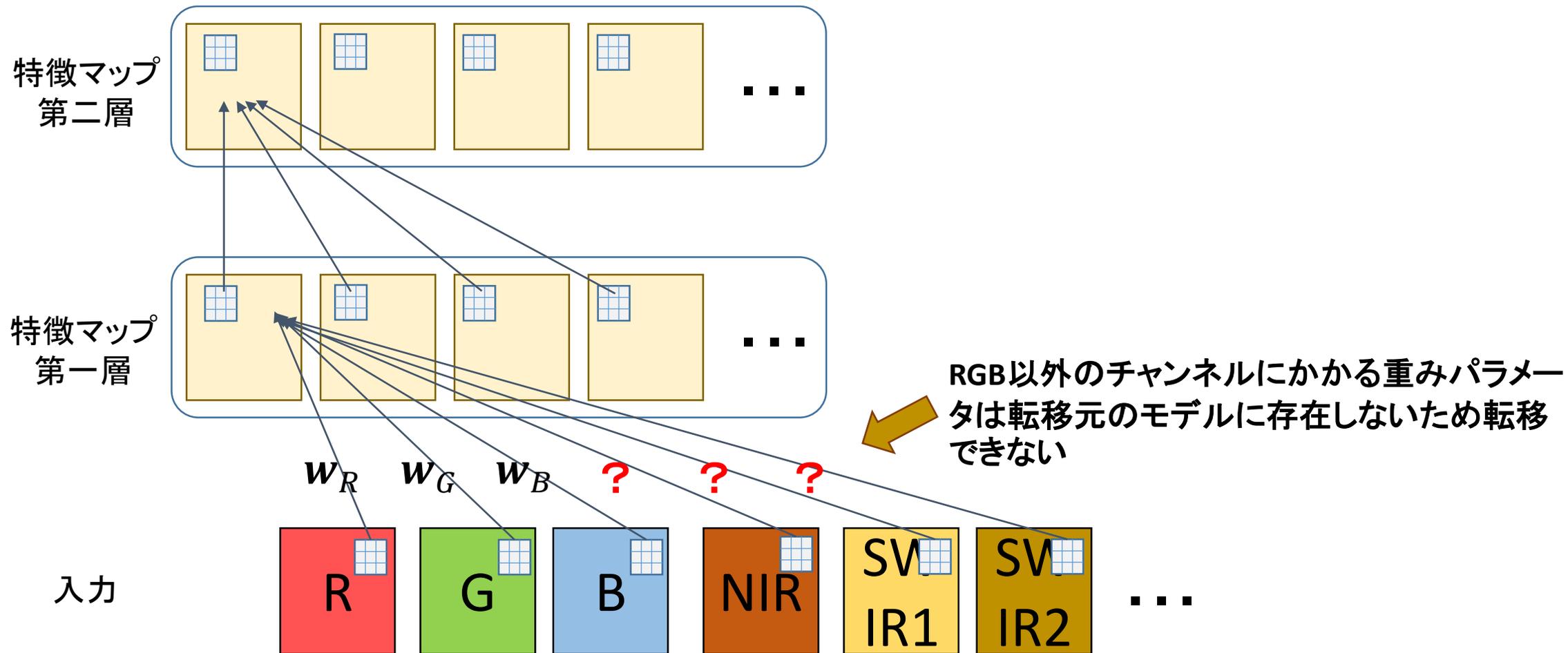


転移学習：マルチスペクトル画像への対処

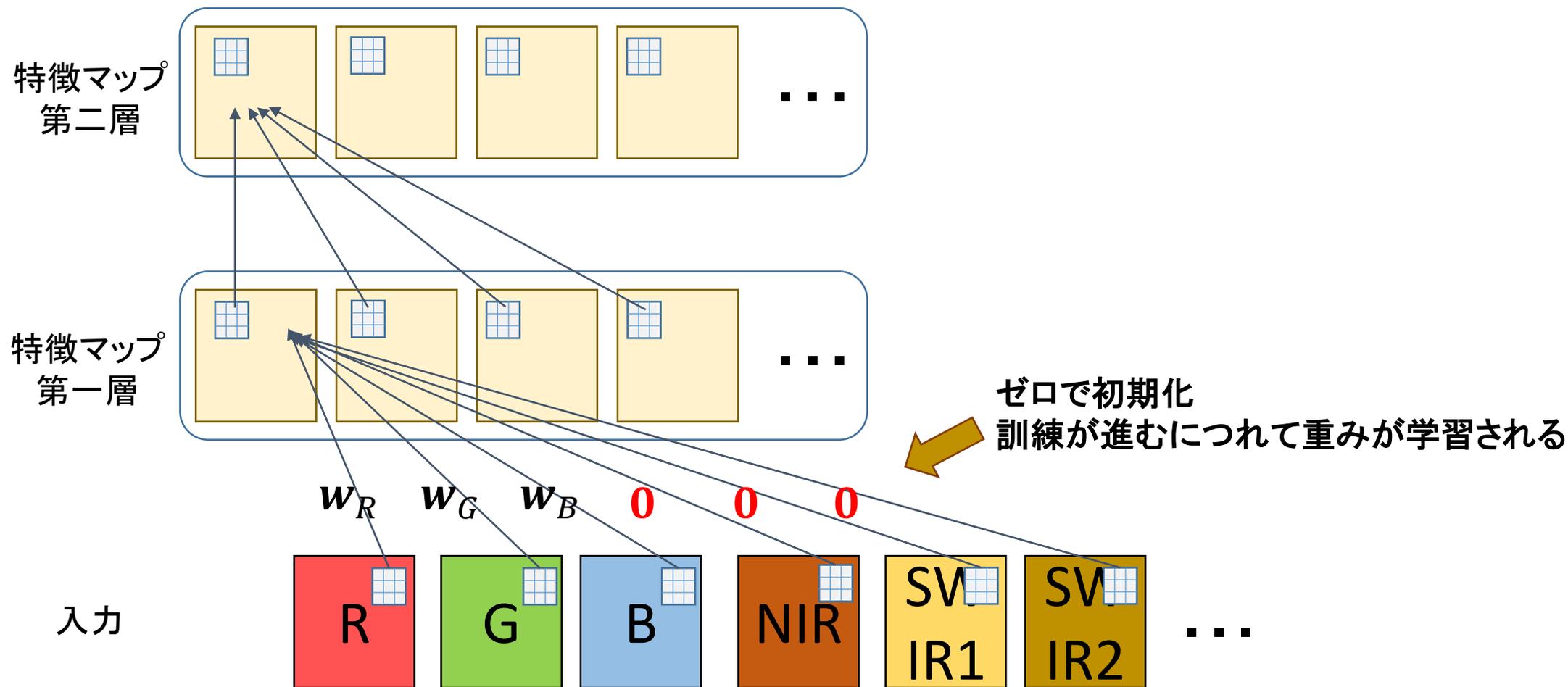


RGBの場合はImageNetの学習済みパラメータをそのまま転移できる

転移学習：マルチスペクトル画像への対処

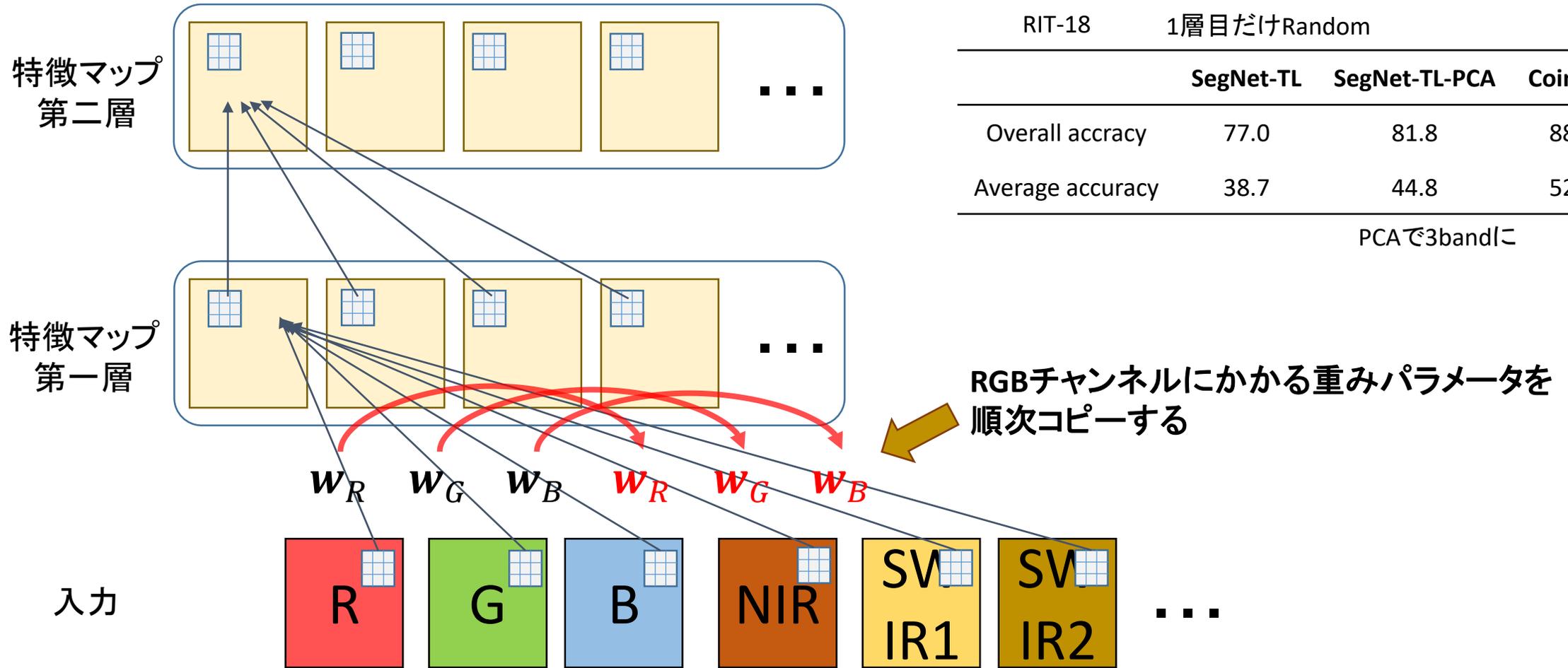


RGB以外のチャンネルの重みはゼロで初期化 [Iglovikov et al., 2018]



Iglovikov, V., Seferbekov, S., Buslaev, A., & Shvets, A. (2018). TerausNetV2: Fully convolutional network for instance segmentation. CVPRW.

同じ重みを他のチャンネルにもコピー [Pan et al., 2019]



RIT-18	1層目だけRandom		
	SegNet-TL	SegNet-TL-PCA	CoinNet
Overall accuracy	77.0	81.8	88.8
Average accuracy	38.7	44.8	52.1

PCAで3bandに

合成画像からの学習 [Kemker et al., 2018]

マルチスペクトル画像に対してモデルを十分に訓練できる教師付きデータがない

→ シミュレーションでマルチスペクトル画像を合成(DIRSIG)

→ 合成画像を使って事前学習



作成した合成画像

精度比較 (Average accuracy)

	ランダム初期値		合成画像で事前学習	
SharpMask	52.4 %	➡	57.3 %	+4.9%
RefineNet	30.1 %	➡	59.8 %	+29.7%

本日の内容

- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - 道路検出はトポロジーまで考える
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

変化検出のターゲット

- 異なる時期に撮影されたリモートセンシング画像を比較して、2時期間の変化を抽出したい
- リモートセンシングで検出すべき変化はどんな変化か？
- リモートセンシングの強みは広域性と継続性
- 広域かつ継続的な観測を活かしてレアな変化をとらえることができれば価値の高いアプリケーションになる
 - 都市の変化（建物の新築・滅失、道路の新設、インフラの新設・取り壊し）
 - 災害の発生（土砂崩れ、道路の寸断、洪水）

レアな変化を検出するタスクの難しさ

① クラスインバランス

② 対象外の変化

Negatives

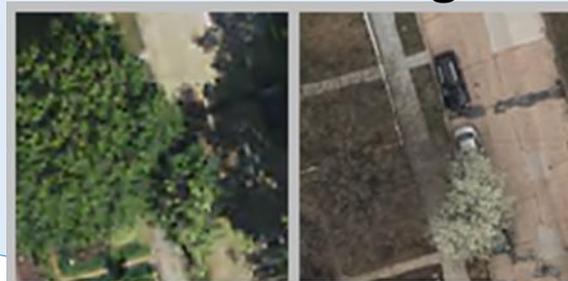
Illumination



Moving objects



Seasonal change



Positives

High-cost

正例が不足する中で対象外の変化にロバストなモデルをいかに構築するか？

コンセプト

- ① クラスインバランス ② 対象外のイベント
- 大量の負例を最大限活用したい**

Negatives

Moving objects

Illumination



Seasonal change



Positives

High-cost

正例が不足する中で対象外のイベントにロバストなモデルをいかに構築するか？

コンセプト

- ① クラスインバランス ② 対象外のイベント
- 大量の負例を最大限活用したい

Negatives

Moving objects

Illumination

負例から抽出できる情報は何か？
⇒ “対象外のイベント”がどういうものかがわかる

Seasonal change

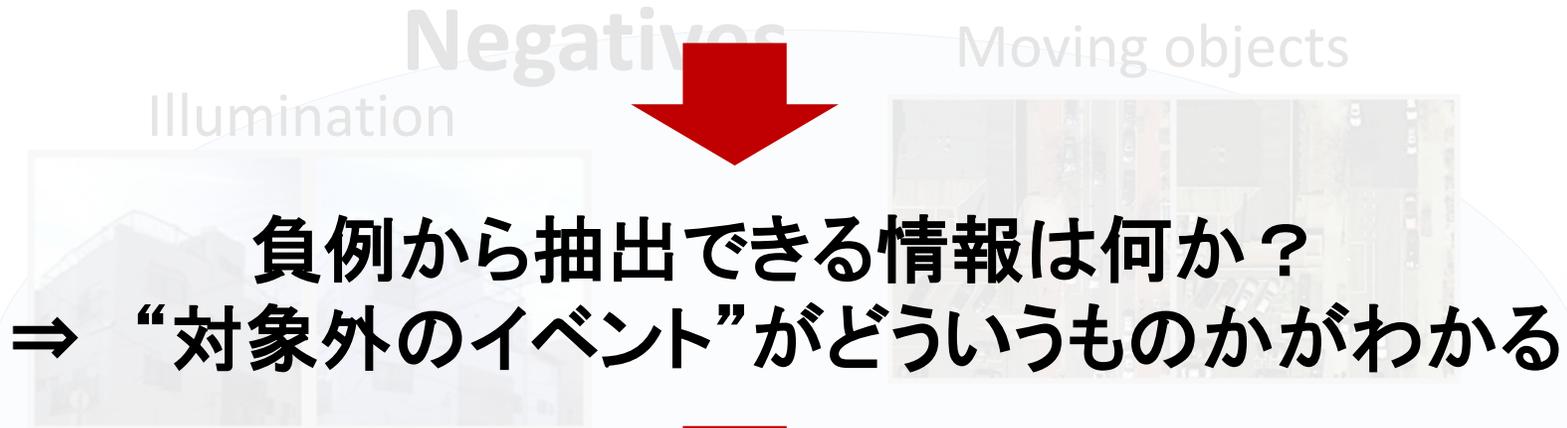
Positives

High-cost

正例が不足する中で対象外のイベントにロバストなモデルをいかに構築するか？

コンセプト

① クラスインバランス ② 対象外のイベント
大量の負例を最大限活用したい



負例から抽出できる情報は何か？

⇒ “対象外のイベント”がどういうものかがわかる

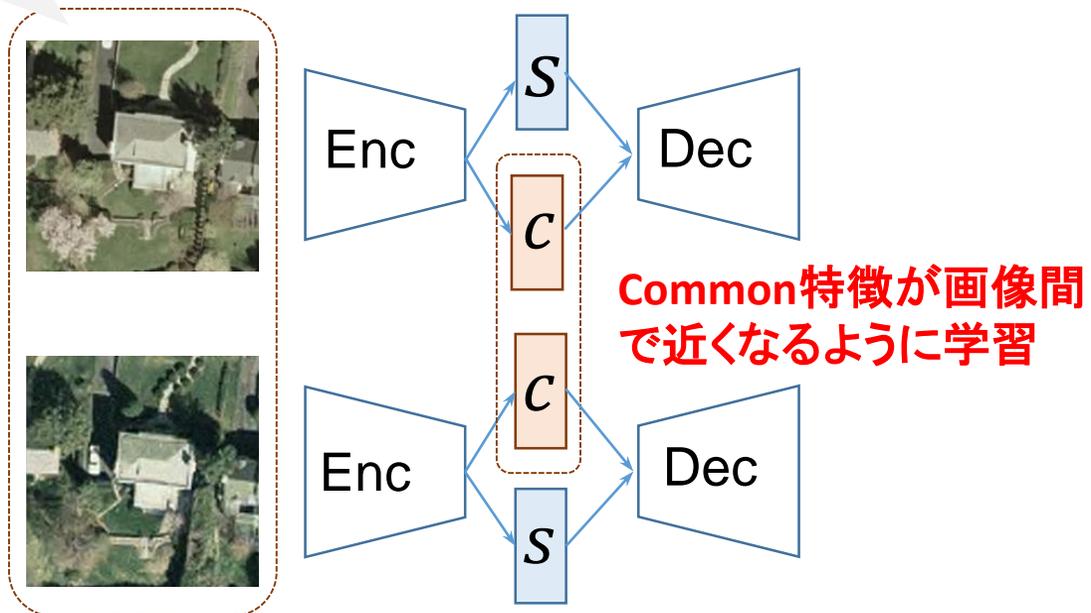
“対象外のイベント”がどういうものか知っていれば
“対象のイベント”を検知するのが容易になる

正例が不足する中で対象外のイベントにロバストなモデルをいかに構築するか？

表現学習を応用した変化検出手法 [Hamaguchi et al., 2019]

STEP1: 表現学習

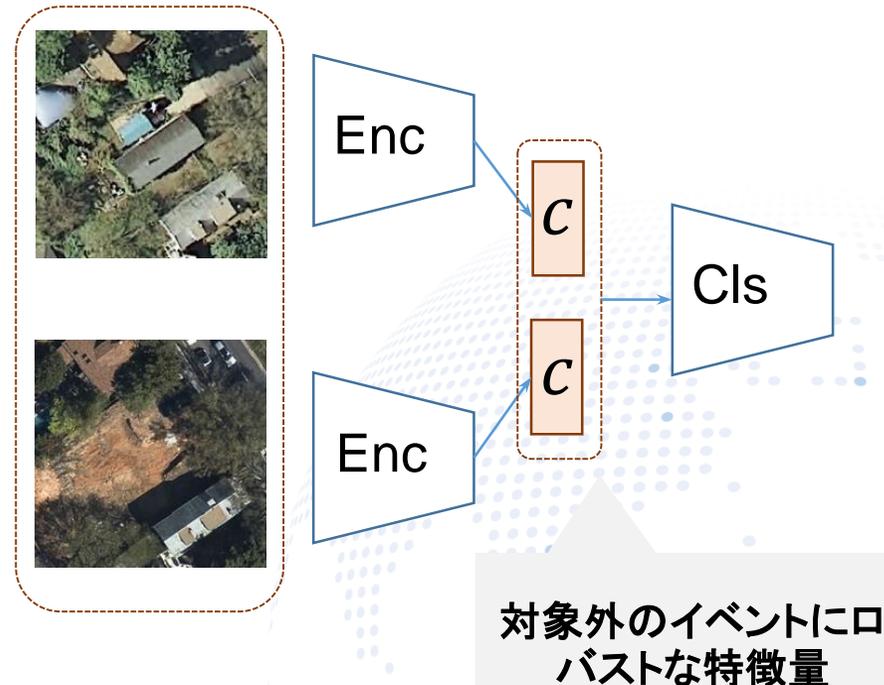
入力は負例のみ



Common特徴: ペアの間で不変な画像コンテンツ
Specific特徴: ペアの間で変化する成分 = 対象外の変化

STEP2: Fine-tuning

入力は負例と少量の正例



本日の内容

- リモートセンシングと深層学習
 - なぜ深層学習か？
 - 深層学習を使って何ができるか？
- リモートセンシング画像の認識は何が難しいか？
 - 解像度の限界を攻める ～小さなオブジェクトを認識するには～
 - 道路検出はトポロジーまで考える
 - ラベル付きデータの不足をどう克服するか？
 - レアな変化ほど価値がある ～変化検出のクラスインバランス問題への対処～
 - 全世界に汎化できるか？ ～大きなクラス内分散への対処～

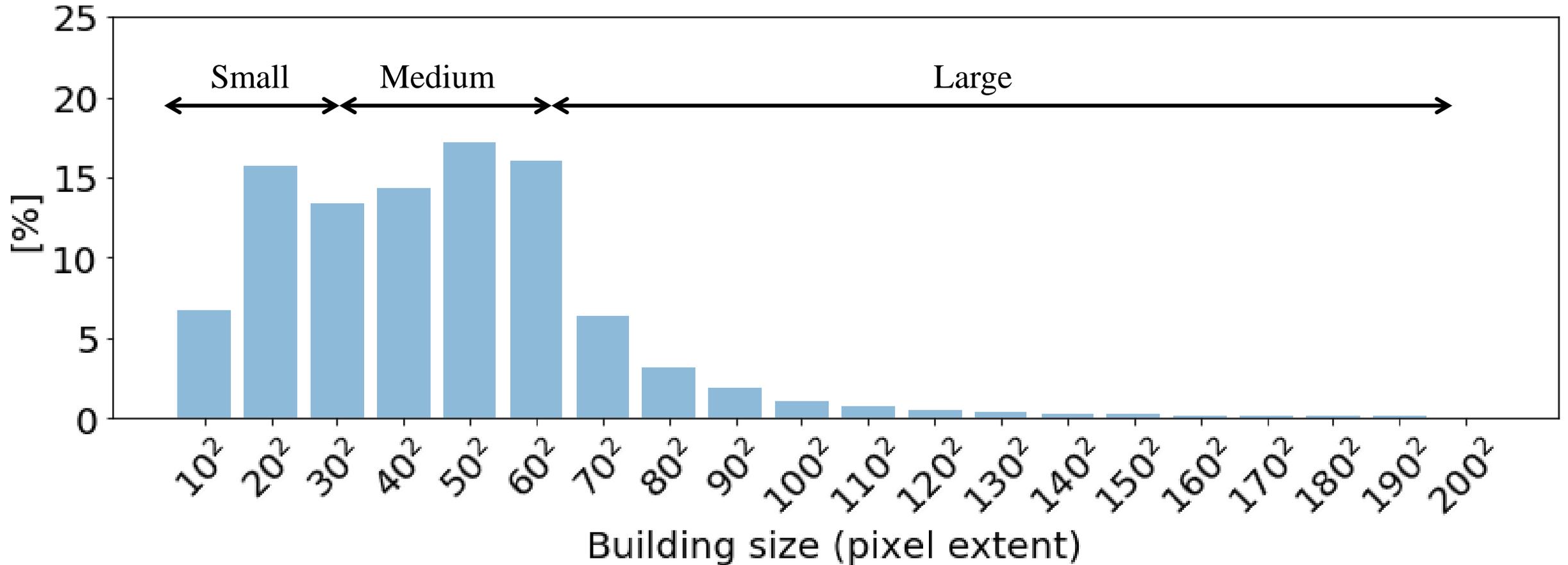
汎化性能を向上させるうえでの障害

- リモートセンシング画像が対象とする建物や道路・農地・樹木といった地物は大きなクラス内分散を持つ
- 例えば建物はサイズによって・地域によって見た目が大きく異なる
- 全世界に適用できるモデルを構築しようとした場合に、非常に大きなクラス内分散を持つデータを扱う必要が出てくる



建物サイズの分布

Distribution of building size in DeepGlobe building detection dataset



様々なサイズの建物の例

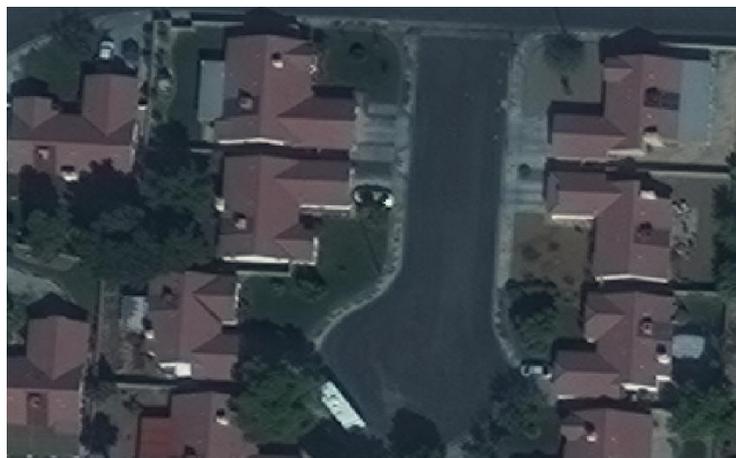
Small



Medium



Large

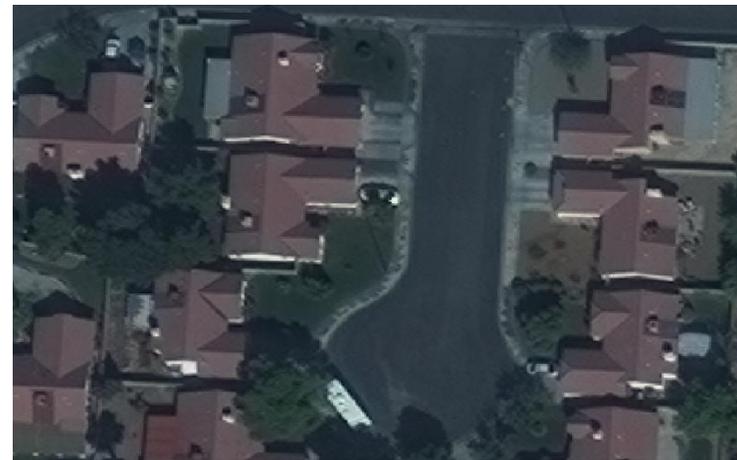


様々なサイズの建物の例

Small



Medium

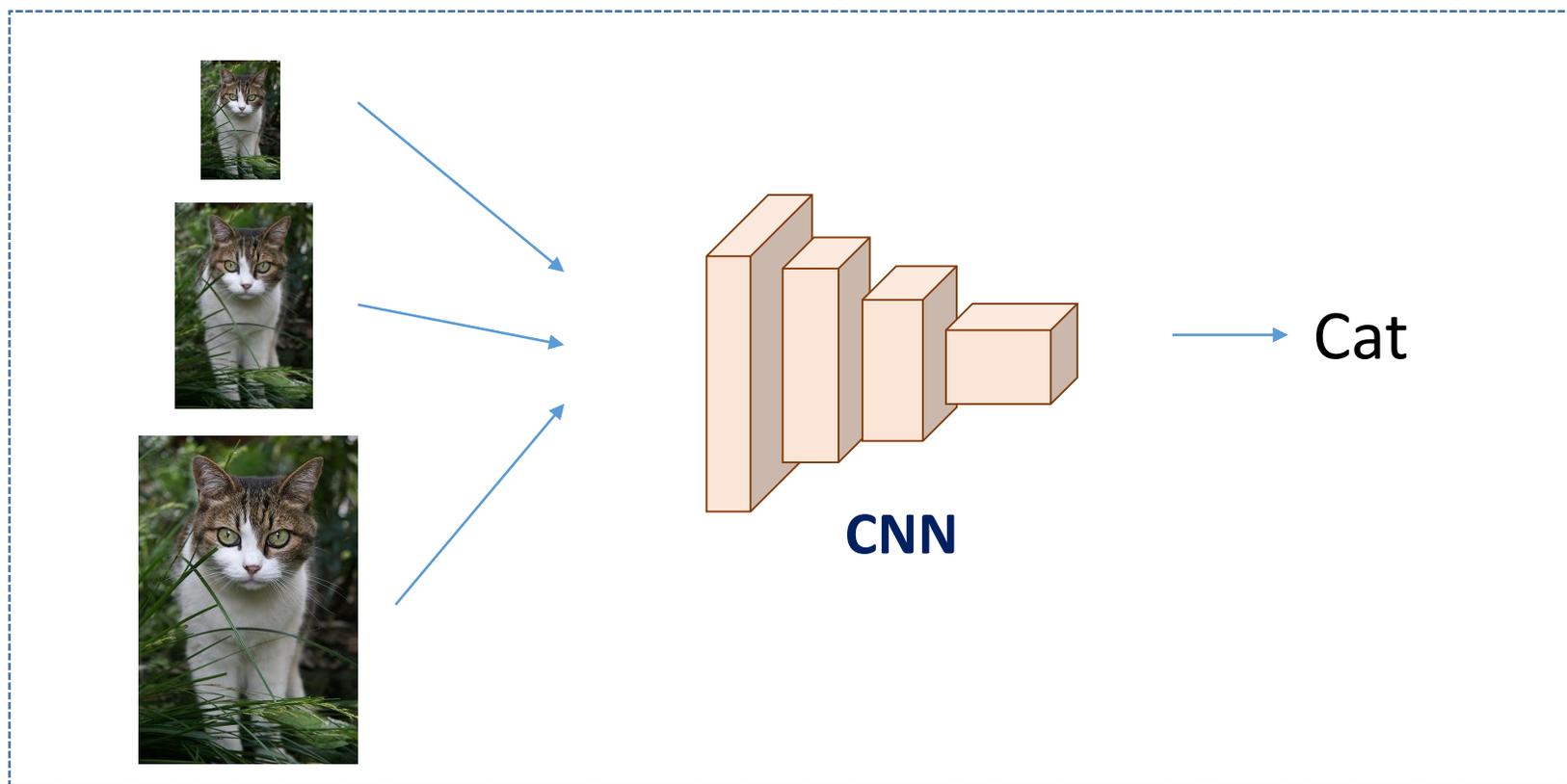


Large



地上視点画像の場合

- サイズのバリエーションは主に対象との距離の違いで発生している
- 拡大・縮小すれば同じような見た目になる
- マルチスケールでの訓練・推論が有効で、実際良く使われる



建物検出の場合

Small



Large



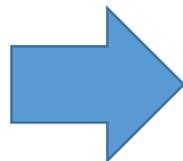
建物検出の場合

- 対象との距離の違いではなく、物体自体のサイズがそもそも違う
- かつ、サイズによって建築様式も異なる → マルチスケールでは対応不可

Small



Up-scaled



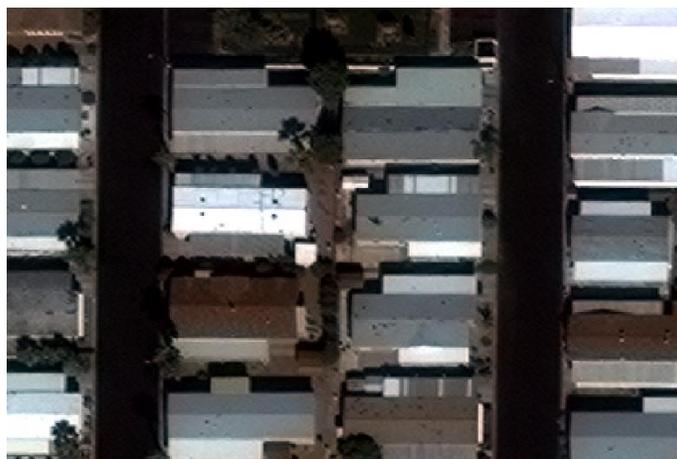
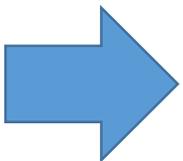
Small (up-scaled)



Large



Up-scaled



サイズ特化型のモデル [Hamaguchi et al., 2018]

