

知っておきたいキーワード

画像認識における転移学習

齋藤 邦章[†]

[†] ボストン大学

"Transfer Learning in Image Recognition" by Kuniaki Saito (Computer Science Department, Boston University, USA)

キーワード：転移学習，事前学習，マルチタスク学習，ドメイン適合

転移学習とは

転移学習 (Transfer Learning) とは、(ある問題を解く際に獲得される) 何らかの知識を、関連している問題を解くのに活用することです。転移学習に明確な定義はありませんが、ある問題を解くのに、そのタスクとは違うタスクのためのデータセットや、そのデータセットで学習されたモデルなどを利用したいというモチベーションがある際によく用いられる用語です (図1)。さまざまな事例が転移学習の一種にあたります。クラス情報などのアノテーションがついたデータを収集するコストをできるだけ抑えたい場合などにデータの得やすい領域の知識を利用しながら、問題を解くことを差す用語とも言えます。

画像認識に関するさまざまなタスクにおいて、Convolutional Neural Networks (CNN)¹⁾ が活用されますが、CNNを学習させる場合にも転移学習を活用することができます。

例えば、ImageNet²⁾ などの大量のラベル付き画像データセットで事前学

習したCNNから得られる特徴量を用いて、自分で収集した小規模な画像データセットに対する識別器を学習させることなどが、挙げられます。大規模なデータセットで学習されたCNNから得られる特徴量は、さまざまなカテゴリーを分類するのに有用な情報を保持しています。このような特徴を利用することで、小規模な画像データ

セットも精度高く識別できることが期待されます。

また、近年、事前学習のみならず、異なるタスク間や異なるドメイン間での知識転移に関して広く研究が行われています。本項では、転移学習に関する近年の研究動向を具体的な事例とともに解説していきたいと思います。

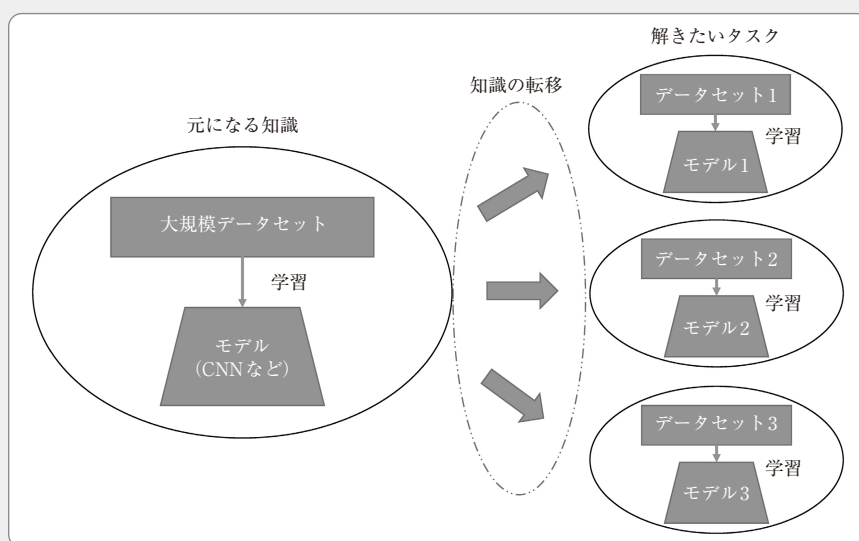


図1 転移学習の概要

事前学習としての転移学習

画像に関するさまざまなタスクを解く際に、ImageNetデータセットで事前学習したConvolutional Neural Networks (CNN) が頻繁に用いられます。ここではその例を取り上げてみたいと思います(図2)。

(1) 物体検出, セマンティックセグメンテーション

1枚の画像を与えられた際に、その画像の属するクラスを識別するのが、Image Classificationにあたります。ある画像の中に存在する物体の位置およびそのラベルを推定するのが、物体検出になります。そして、画素単位で物体のクラスを推定するのが、セマンティックセグメンテーションになります。これらのタスクのためにネットワークを学習させる際には、ImageNetで学習されたCNNをFine-tuning (事後学習) することがほとんどです。ImageNetのように大規模なデータセットで学習されたCNNは、さまざまなクラスの物体を識別するのに効果的な特徴を保持すると考えられ、実際にその効果が各タスクで確認されています。

(2) Image Captioning, VQA

画像を説明する説明文を生成するタ

スクがImage Captioningにあたります。それに対して、画像とその画像に関する質問文が与えられた上で、正しい答えを返すようなモデルを学習させるのが、VQAに当たります。これらのタスクにおいても、画像の特微量には、ImageNetでの学習で得られたCNNを用いることがほとんどです。このよう

に、画像の認識のみならず、画像と言語の関連を考慮するようなタスクにおいても、ImageNetから得られる知識を転移させることが、精度高くタスクを解く鍵となっています。これは、事前学習されたCNNが画像内に存在する物体のクラスに関するリッチな情報を保持しているためと言えます。

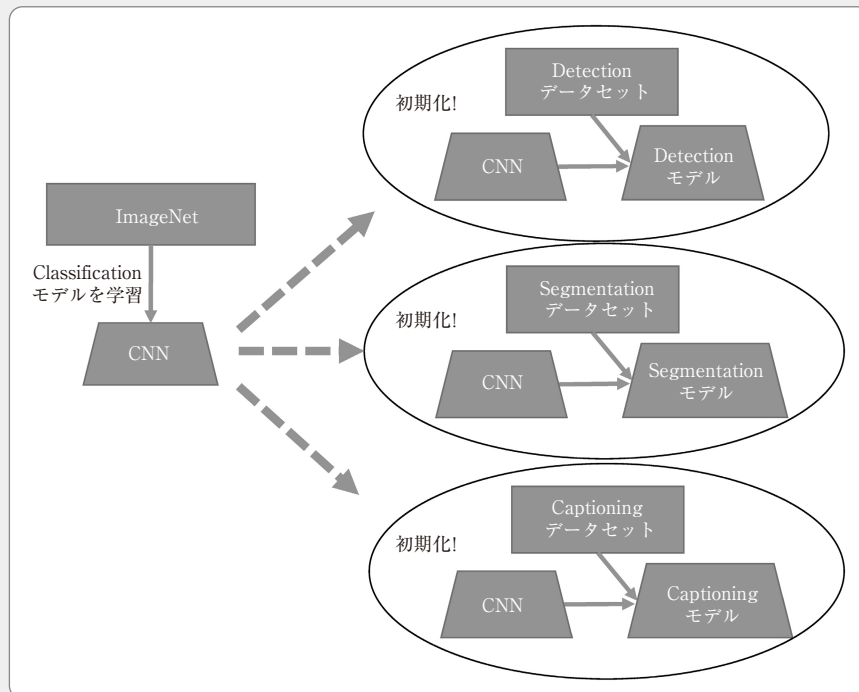


図2 Detection, Segmentation, Captioning における転移学習

マルチタスク学習

異なる複数のタスクが存在する際に、それぞれのタスクが、どのように関連しているのかを考慮し、異なるタスクから得られる知識を利用することで、それぞれのタスクを精度高く解くというのが、マルチタスク学習になります。例えば、先に挙げた物体検出とセマンティックセグメンテーションは、直感的には関連の深いタスクであると言えます。どちらのタスクにおい

ても、画像内のある位置に対してどんな物体が存在するのか、またはしないのかを判断する必要があります。その一方で、セグメンテーションにおいては、物体の形状をピクセル単位で推定する必要があるのに対して、物体検出では、物体を囲む矩形領域を出力するのみで良いという違いもあります。効率よく知識の転移を行うには、どのような情報がタスク不変に有用で、どのような情報がタスク固有に有用な情報なのかということを考慮する必要があ

るということになります。Kendallら³⁾は、図3に示すマルチタスク学習のためのネットワークを使用しました。三つのタスクに対して、共有されたネットワークと、それぞれのタスクに対する出力を行うためのネットワークを使用しています。どのタスクを優先的に学習するかという重みについては、人間が何らかのヒューリスティクスに基づいて決定するということが考えられますが、この研究においては、

それぞれのタスクに関するネットワークの確信度というものに基づいてタスクに対する重みそのものを学習する方法を提案しています。Computer

Visionにおいては、非常に多くのタスクが存在しますが、2018年に開催されたComputer Visionのトップ会議の一つであるCVPRにおいては、さまざま

なタスクの関連度を調べるような研究がBest Paper賞を受賞しました⁴⁾。転移学習に対する注目度の高まりを感じるところです。

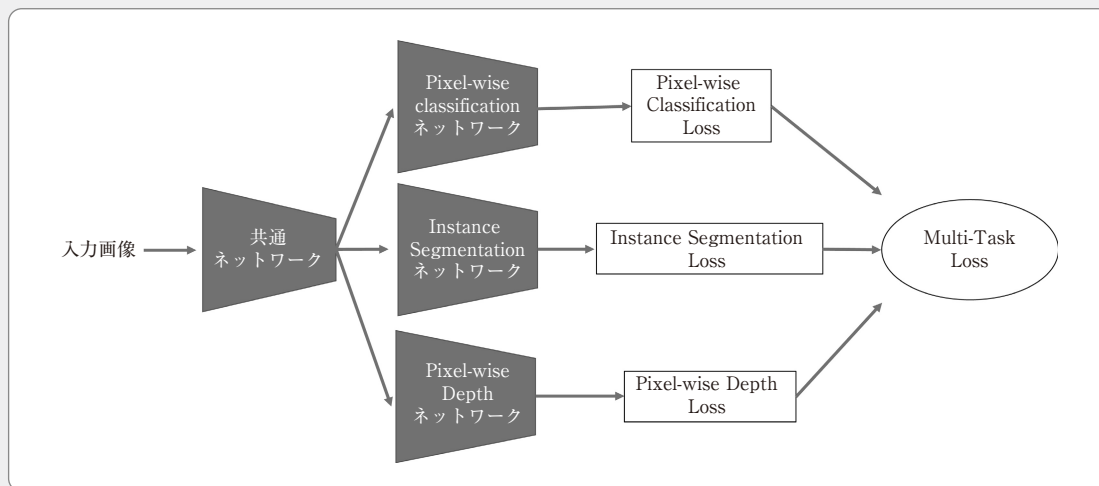


図3 マルチタスク学習の例

ドメイン適合

同じタスクに対するデータセットではあるが、異なるドメインで収集されたデータセット同士でいかに知識を転移させるかを考慮するのが、ドメイン適合 (Domain Adaptation) になります。その際、知識の元となるドメインをソースドメインと呼びます。そして、実際に精度高く認識を行いたいドメインをターゲットと呼びます。例えば、最近の研究においては、自動運転用の

シミュレーション環境から得られた画像 (ソース) を実際の車載カメラで得られる画像 (ターゲット) の認識に使うという研究が多く行われています (図4)。シミュレーション環境内には3Dモデルが事前に定義してあるので、得られる画像内のどの部分に何の物体が存在するのか、などの情報を容易に得ることができます。そのようなデータから得られる知識を、さまざまな環境が存在する実世界データの認識に活用しようというものです。しかしなが

ら、シミュレーション環境と実世界で得られる画像には、画像全体のテクスチャや、物体の色などに違いが見られます。そういったドメインの差を考慮しながら、ターゲットにおいて高い精度を示すモデルを学習させる手法が数多く提案されています (図4)。ドメイン適合の応用先は自動運転にとどまりませんが、このタスクに対して非常に多くの手法が近年のComputer Vision学会において提案されており、注目度の高さを感じます⁵⁾。

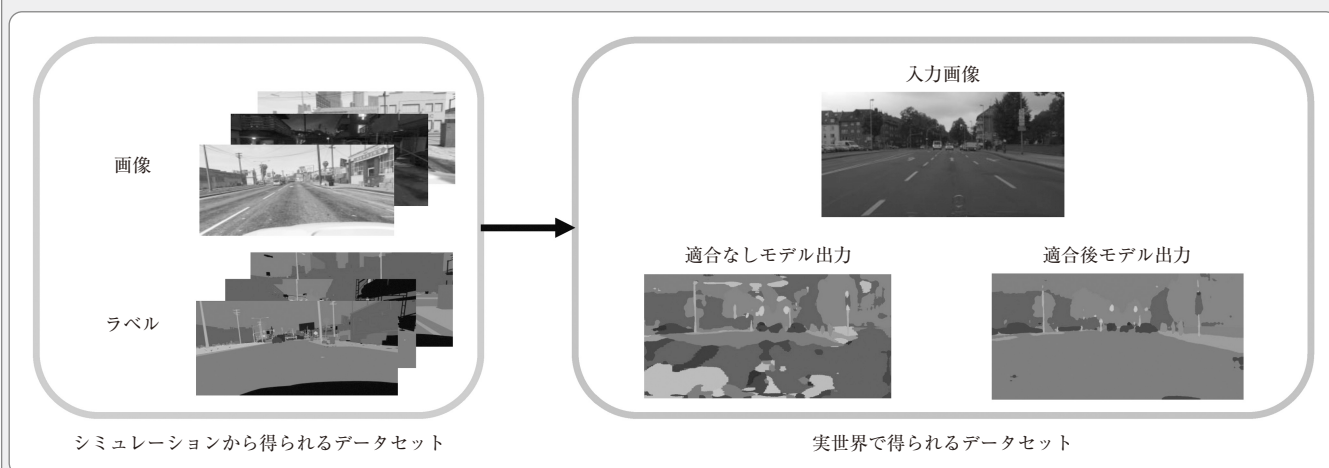


図4 Semantic Segmentationにおけるドメイン適合の例

むすびと今後

今回のコラムでは、画像認識における転移学習の代表的な事例を紹介しました。転移学習はデータを集めるのが困難な状況に対して、有効な学習手法であり、応用先は、今回紹介した事例

にとどまりません。言語コーパスから得られる情報を見たことのない画像を認識するのに用いる Zero-Shot 学習⁶⁾など、非常に興味深い事例も多く存在します。また、大規模なデータを用いた場合には、既存のモデルで非常に高い認識精度を達成することができます

が、アノテーション付きのデータを多く得られないような状況においては、高い精度を達成することが既存の手法ではまだ困難です。とは言え、多くの研究者がこのトピックに関心を持っており、今後も研究が着実に進んでいくと感じています。 (2018年10月9日受付)

参 考 文 献

- 11) Deng, Jia, et al.: "Imagenet: A large-scale hierarchical image database", Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE (2009)
- 2) Krizhevsky, Alex, I. Sutskever and G.E. Hinton: "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems (2012)
- 3) Kendall, Alex, Y. Gal and R. Cipolla: "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
- 4) Zamir, R. Amir, et al.: "Taskonomy: Disentangling Task Transfer Learning", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
- 5) Saito, Kuniaki, et al.: "Maximum classifier discrepancy for unsupervised domain adaptation", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
- 6) Socher, Richard, et al.: "Zero-shot learning through cross-modal transfer", Advances in neural information processing systems (2013)



さいとう くにあき
齋藤 邦章

2018年、東京大学
大学院情報理工学
研究科修士課程修
了。同年より、米
ボストン大学
Computer Science

Department PhD Programに入学し、
現在に至る。

キーワード募集中

この企画で解説して欲しいキーワードを会員の皆様から募集します。ホームページ (<http://www.ite.or.jp>) の会員の声より入力可能です。また電子メール (ite@ite.or.jp)、FAX (03-3432-4675) 等でも受け付けますので、是非、編集部までお寄せください。 (編集委員会)