［研究助成（A）］

画像空間の構造と画像変換ネットワークの構造の関係の研究

Relationship between the structures of image spaces and image transformation networks

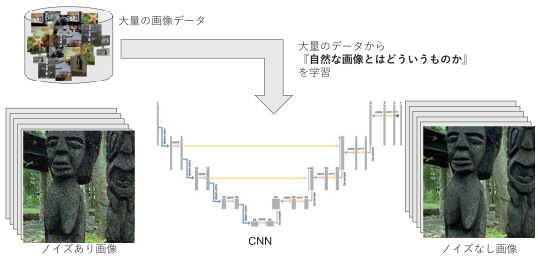
2201004

研究代表者　早稲田大学　理工学術院
基幹理工学研究科　教　授　石　川　博

［研 究 の 目 的］

データ空間が構造を持つということは，データを格納する変数が意味を持つことである。プログラム言語理論では，プログラムの効果・意味をセマンティクスと呼ぶが，ここでは，いわば変数のセマンティクスを考える。つまり，計算機応用で通常採用される「符号化は任意である」という概念とは対照的に，データは様々な構造を持つ空間の要素であり，それら空間の間には構造について自然な関係があると仮定する。例えば，画像は平面領域の各点に色を与えたものなので，平面領域の構造と色の空間の構造を受け継いでいる。データの，そしてそれから誘導されるデータ空間の構造を考慮することによって，その視点から様々なタスクにおけるネットワーク構造とその汎化性能の関係を理解することをめざす。

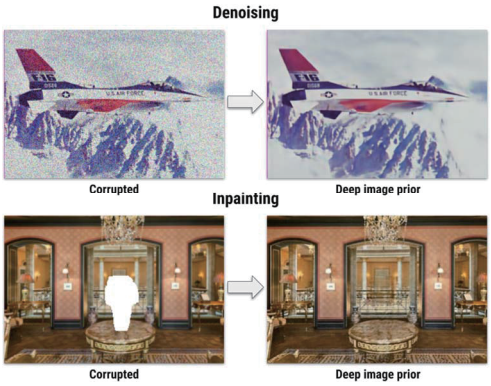
一般のニューラルネットワークは実数値をとるという以外のデータ空間の構造を考慮していないが，コンピュータービジョンに劇的な性能向上をもたらした畳み込みニューラルネットワーク（CNN）は，一般ネットワークの構造に加えて，変数に空間的な意味を与え，画像領域における近傍や対称性の概念をその構造に反映させている。このことがその性能の元であるという証拠の一つは Deep Image Prior［Ulyanov et al. CVPR2018］である。これは学習によらず最適化によって，ネットワークの表す関数が自然に汎化性能を持つようになることを示した。



一般の深層学習

汎化こそが機械学習の本質である。これを説明するため例えば画像のノイズ除去を考えると，一般の深層学習では，学習データを予測するようにモデルのパラメータ（重み）を決定する。つまり，大量の（ノイズ入り入力，ノイズ無し正解出力）組からなる学習データを与えて，入力に対するネットワークの出力と正解出力の距離が小さくなるような重みを探す。

つまり，ここでいう学習とは，入力に対する出力を与える関数を覚えさせることであるが，単なる記憶とは重要な違いが存在する。それは，記憶とは前に見た学習データに対して入力ー出力関係を答えられることであるが，学習における目標は，学習データに入っていない，初めて見る入力に対する出力を「推測」することであるということである。これを汎化という。つまり，学習データで見たことをおしひろげて，まだ見ぬデータにおける入力-出力関係を推測するのである。問題は，何を根拠に推測するかで



（上段：ノイズ除去　下段：画像補完）

Deep Image Prior

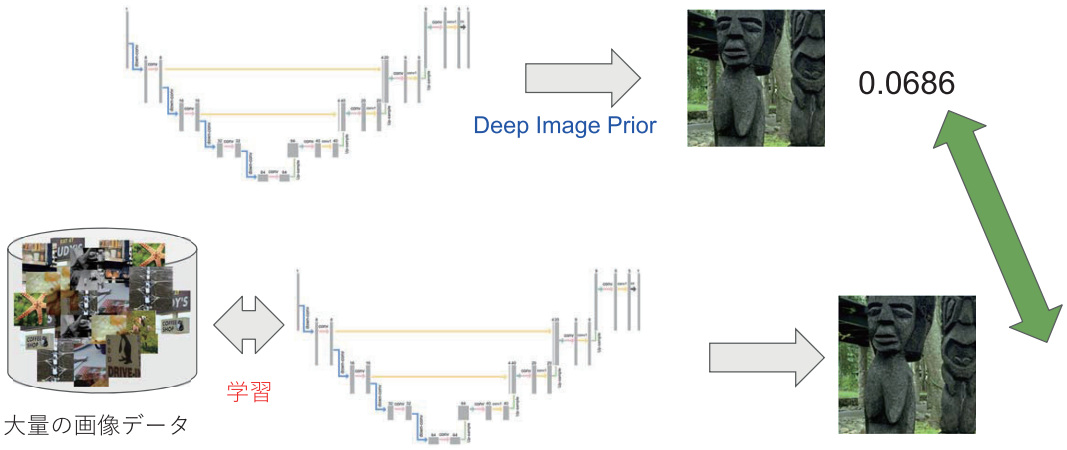
ある。

汎化が可能なのは，学習データの入力と出力の間に何らかの相関関係が存在して，それをデータから「学習」するからだと考えられる。ところが，Deep Image Prior では，学習データを全く使わず，ランダムな入力を固定し，与えられたノイズ入り画像にネットワーク出力画像が近づくように，重みだけを変えていく。CNN のパラメータ数は一般に非常に多いため，任意の画像を出力するに十分な自由度を持っていて，最終的には Deep Image Prior CNN の出力は，最適化の目標としたノイズ入り画像に到達する。Deep Image Prior の発見は，重みを変えてそこへ行く途中で，ネットワーク出力画像がノイズを除去した画像になったことである。

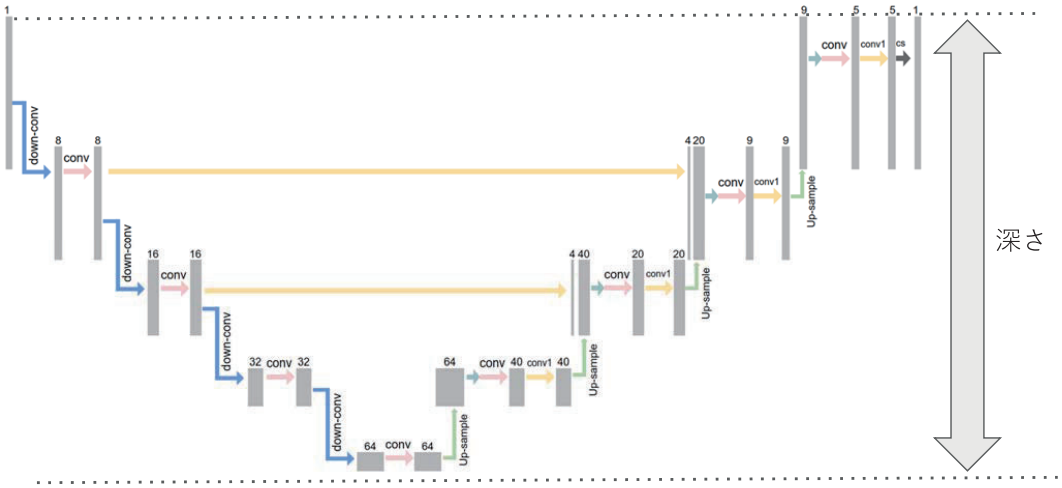
CNN にノイズとは何かを全く教えなくても，出力をノイズ入り画像に近づけていくと，ノイズを除去した画像を途中で通るのである。これは何を意味するのか。まるで CNN が自然な画像を好むかのようである。

ノイズ除去だけであればそれはそれほど特筆すべきことに思われないかも知れない。ノイズ除去はニューラルネットワークによらなくとも，様々な手法によって学習データなしに達成することができるのだから。しかし， Deep Image Prior はノイズ除去だけでなく高精度の画像補完（inpainting）も達成した（左図）。学習なしに純粋に CNN の重みを一定の方法で最適化するだけでここまで自然に画像の穴を埋めることができるのは完全に予想外であった。Deep Image Prior においては全く学習させないのであるから，CNN のネットワーク構造自体がこれを引き起こしているように思われる。

このように Deep Image Prior は，学習による重みだけでなく，ネットワークの構造が汎化性能に重要であることを示したが，これは当然，同じネットワークで学習をする場合にも重要であると思われる。そこでこの二つの関係を多様な画像変換問題において調べ，より具体的にどの要素が重要であるかを解明することにより，各タスクにおける，より高性能な学習システムの開発につなげることをめざす。



同じネットワークの２つの使用法



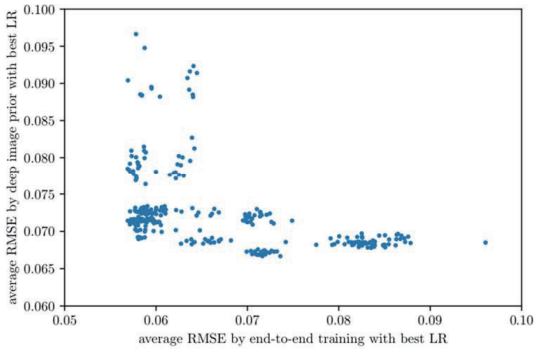
構造の「深さ」に着目する

Encoder-Decoder アーキテクチャ

［研究の内容，成果］

本研究では，上記の手始めに，同一のネットワーク構造の，通常の学習によるタスクにおける性能と，Deep Image Prior としての学習をしないタスク性能との相関を調べた。これは，Deep Image Prior としての性能の方が学習の必要がないため短時間で測定でき，もし両者に相関があれば，学習により使用する CNN の構造設計をより高速に行えるためである。そのために，ノイズ除去タスクにおいての両者の性能を調べた。

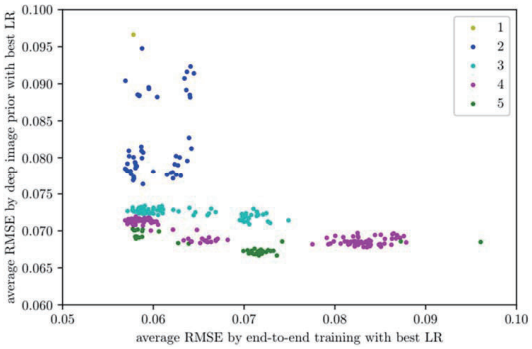
調査した構造は，いわゆる Encoder-Decoder アーキテクチャ（上図）と呼ばれるもので，前半の Encoder 部分が入力画像から畳み込み層により解像度を下げつつチャンネル数を増やしていき，画像の特徴を抽出する。そして，図で一番低い部分にあたるボトルネックを通して，今度は Decoder が画像に戻す。この基本構造に，スキップコネクション（SC）と呼ばれる，ボトルネックをバイパスする経路を付加する。これら Encoder や Decoder の深さや，SC の有無，また SC を入れる層の深さの違いによる様々な構造 300 種類以上について調査した。



学習時と Deep Image Prior 時の性能プロット

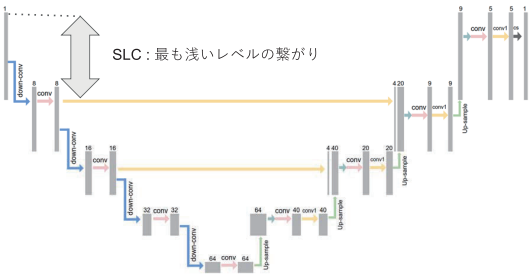
それぞれの構造について，ノイズ除去タスクのための通常の学習をした場合の最終的なテスト評価値 RMSE と，学習をしない Deep Image Prior としての評価値とをプロットしたものが上図である。横軸が学習時の性能，縦軸が Deep Image Prior 時の性能である（以下同様）。これを見ると，両者には何らかの相関があるように見える。それぞれの場合に何が性能差をもたらしているかが問題になる。

種々の可能性を検討したが，結論としては，ネットの深さ（上図参照）により色分けした次のプロットを見ると，同じ色が横に並んでいることから，Deep Image Prior 時の性能（縦軸）は，ネットワークの深さに強く依存していることがわかる。



深さによる色分け（右上が凡例）

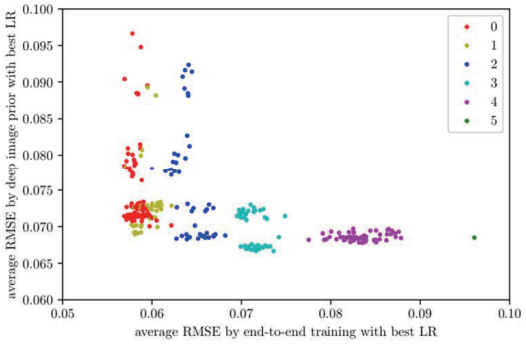
一方，同様なことが学習時の性能（横軸）について見つからずに苦労したが，最終的に，次のようなことが分かった。



「デコーダ側への、最も浅い繋がり」に着目

SLC

SLC（Shallowest Level of Connection）とでも呼ぶべき量に注目する。これは，スキップコネクションのうち，最も浅いレベル間をつなぐものの深さである。これによって色分けすると，プロットは次のようになる。



SLC による色分け（右上が凡例）

今度は，同じ色が縦に並んでおり，横軸，つまり学習時の性能はこの SLC という量に依存しているようである。

以上のように，学習によるタスクの性能は主に，最も浅い SC の深さに関係し，Deep Image Prior としての学習をしないタスク性能は，ネットワークの深さに関係することが判明した。これらの 2 要素は独立に変化させることができ，実際にこれらを変化させると 2 つの方法によるタスク性能をコントロールすることができるので，当初の目論見である CNN 構造設計には単純には使えないことが判明した。これとともに，プロットが一見相関を表すように見える理由も判明した。すなわち，ネット全体が浅い場合には，SLC も大きくはできないということである。そのため，プロットの右上部分，つまり深さが浅く SLC が大きい場合に対応する部分にはプロットの点が存在しない。

［今後の研究の方向，課題］

本研究の当初の目論見は外れたが，今後もデータの，そしてそれから誘導されるデータ空間の構造を考慮することによって，その視点から様々なタスクにおけるネットワーク構造とその汎化性能の関係を理解することをめざしたい。

［成果の発表，論文等］

検討中