

# Is the deconvolution layer the same as a convolutional layer?

CVPR 2016の論文[1]では、単一画像のスーパーレゾリューション (SR) を行うための新しいネットワークアーキテクチャを提案しました。既存の畳み込みニューラルネットワーク (CNN) ベースのスーパーレゾリューション手法[10,11]のほとんどは、まずバイキュービック補間を用いて画像をアップサンプリングし、その後、畳み込みネットワークを適用するものである。ここでは、まず画像をアップサンプリングすることから、この種のネットワークをハイレゾリューション (HR) ネットワークと呼ぶことにする。その代わりに、図1のように、低解像度 (LR) の入力を直接サブピクセルCNNに与えます。

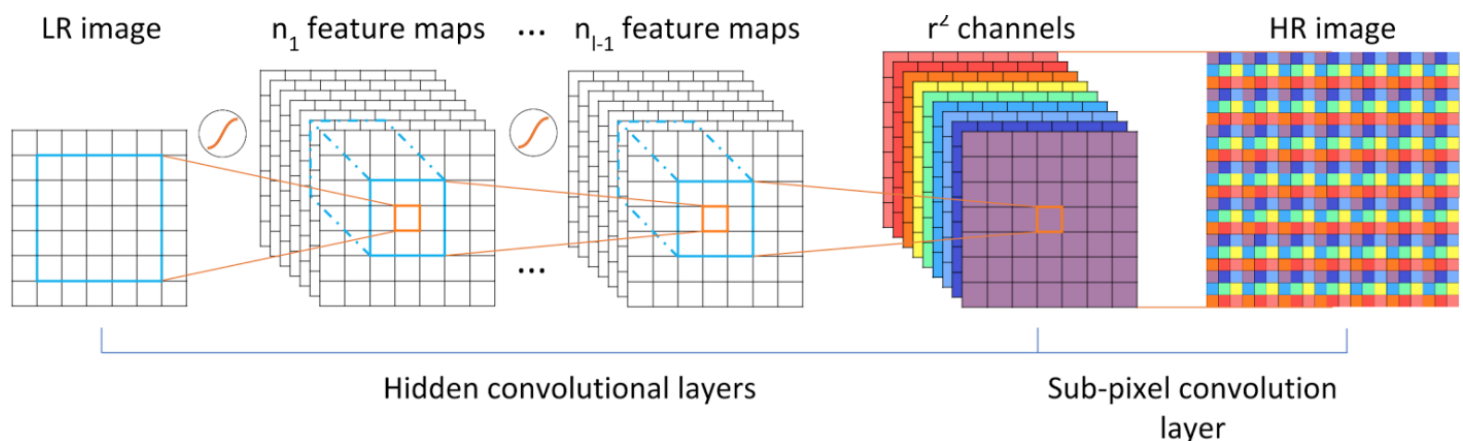


図1：ESPCNフレームワークの説明図 ( $r$ はアップスケーリング率を表す)。

例えば、入力のLR画像が  $1 \times 1$  であれば、出力のHR画像は  $r \times r$  となります。このように、1枚の高解像度 (HR) 画像の代わりに  $r$  個のチャンネルを出力し、周期的な2回のシャフリングを行ってHR画像を再構成します。この効率的なサブピクセル畳み込み層の動作についての詳細は、論文に記載されています。ここでは、このネットワークをLRネットワークと呼ぶことにします。

このノートでは、このネットワークを見たときに CVPR でほとんどの人が質問した2つの質問に関連する2つの側面に焦点を当てたいと思います。第一に、なぜ  $r$  チャンネルが魔法のようにHR画像になるのか？2つ目は、なぜLR空間での畳み込みがより良い選択なのか？これらは、論文の中で私たちが答えようとした重要な質問ですが、ページ数の制限があったため、思ったほど深く、明確に説明することができませんでした。これらの疑問に答えるために、まず、転置型コンボリューション層、サブ

ピクセルコンボリューション層、そして我々の効率的なサブピクセルコンボリューション層の関係について、第1章と第2章で説明します。一般的なsubpixel convolutional layer [5]と区別するために、我々の効率的なsubpixel convolutional layerをLR空間のconvolutional layerと呼ぶことにします。そして、一定の計算量と複雑さの場合、LR空間のみで畳み込みを行うネットワークは、最初にHR空間で入力をアップサンプリングするネットワークよりも、同じ速度でより多くの表現力を持つことを示す。