

深層学習ネットワークによるひび割れ検出精度の違い Differences in crack detection accuracy due to some deep neural networks

天方匡純*¹ 吉田龍人*¹ 藤井純一郎*¹ 高橋悠太*¹ 大久保順一*¹
Masazumi Amakata Ryuto Yoshida Junichiro Fujii Yuta Takahashi Junichi Okubo

*¹ 八千代エンジニアリング株式会社
Yachiyo Engineering Co., Ltd.

Abstract: The concrete structures are applied to various social asset facilities like roads, bridges, banks, tunnels, and sewers and support our lives. We will enter the maintenance era and the population decreasing era, and these structures are required as useful inspections and assessments as possible. Image processing is expected as a way of efficient inspections. Especially crack detection with the deep neural network is often seen as examples of Poc. But those examples are limited to each field application and don't lead us to the technical system which social assets technician, due to its nature must establish. The Deep neural network truly improved the generalization of image processing. When we apply their technologies to each industry, we must try our technical adjustments and ingenuities like network structures and parameters. Now I will focus on semantic segmentation which we often use in crack detection. I will uncover differences between SegNet, U-Net, and FusionNet, which we select as a semantic segmentation method, and refer to crack detection accuracy due to the function. As a result, I will organize a part of their knowledge leading to the civil engineering field's technology system.

1. はじめに

コンクリート構造物は、道路、橋梁、堤防、トンネル、下水道など様々な社会資本施設に取り込まれ、我々の生活を支えている。そして、社会資本施設も維持管理の時代、人口減少時代を迎え、各施設の管理者は機能維持のために可能な限り効率的な点検・評価を目指している。点検効率の一つの手段として画像処理の研究開発や実用試行が進むなか、深層学習ネットワークを用いたコンクリート構造物のひび割れ検出は良く見る事例である。しかし、その事例は各自適用に留まり、国民生活を支える社会資本の性質上、不可欠となる技術体系化には至っていない。深層学習ネットワークにより画像処理の汎用性は大幅に向上したが、それでも各産業での現場適用に際しては、各産業独自のノウハウを反映した技術的工夫・調整が必要である。

そこで、社会資本分野、土木分野のひび割れ検出に用いられる semantic segmentation 技術に着目し、その際に主に用いられる深層学習ネットワークである SegNet[Badrinarayanan 17], U-Net[Ronneberger 15], FusionNet[Quan 17]の機能上の違いを明らかとし、これを要因とするひび割れ検出精度の違いに言及し、土木分野への適用を踏まえた際の技術体系化に繋がる一知見を整理する。

2. ひび割れ検出の Semantic Segmentation

社会資本施設のひび割れ等検出とは図 1 のようなものであり、社会資本施設に関する現行の点検要領等への対応からひび割れの幅や長さが検出可能な Semantic Segmentation のタスク

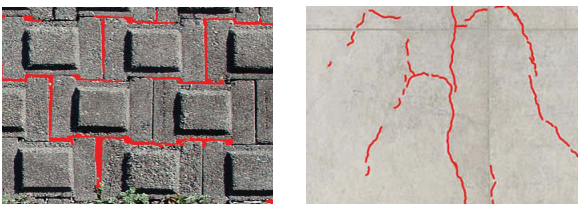


図 1 コンクリート構造物のひび割れ検出の例

連絡先: 天方匡純, 技術創発研究所(RIIPS), 〒111-8648 東京都台東区浅草橋 5-20-8, amakata@yachiyo-eng.co.jp

が選定されることが多い。そこで、本章ではその代表的な深層学習ネットワークの概要を整理する。

2.1 SegNet

2017 年にイギリスの Cambridge 大学のチームが提案した Encoder-Decoder 型の semantic segmentation ネットワークである (図 3 参照)。Encoder の構造は 2014 年に同大学のチームが提案した VGG16 をベースとしている。そして、Decoder は Encoder と対となる逆方向の upsampling を目的とした構造となっている。Encoder の pooling 情報を Decoder に伝えて出力画像の位置再現精度を高める pooling indices が特徴的である。自動運転自動車での利用を想定した道路空間の semantic segmentation の利用例が有名である。

2.2 U-Net

2015 年にドイツの Freiburg 大学のチームが提案した Encoder-Decoder 型の semantic segmentation ネットワークである (図 4 参照)。最初に顕微鏡画像の神経構造や細胞構造の検出に適用されたが、現在では多方面で多活用されている。入力情報を圧縮する Encoder 部分と圧縮情報を拡張する Decoder 部分は、Conv(Decoder では upConv), Max-pooling からなる。特に Encoder 情報を Decoder に伝える連結(concatenate)により圧縮情報から出力情報の過程で位置再現精度を高めていることが特徴的である。

2.3 FusionNet

2017 年にアメリカの Harvard 大学のチームが提案した Encoder-Decoder 型の semantic segmentation ネットワークである (図 5 参照)。U-Net と同じ医療系の顕微鏡画像からの脳の細胞膜や細胞体の検出、そして、細胞形態の統計解析等の活用例で論文紹介されている。Encoder も Decoder も Conv(Decoder では upConv), Max-pooling, Residual block (short skip connection 含む) の組合せで U-Net と基本構造が類似している。大きな違いは、Encoder と Decoder を繋ぐ long skip connection が U-Net の concatenate 構造ではなく、residual 構造である所である。この結果、deep network の問題となる勾配消失問題を解消し、学習効果の高いネットワークとなっている。

3. 特徴量の把握

紹介した 3 つのネットワークは何れも Encoder-Decoder 型の Semantic Segmentation モデルである。Encoder 区間で高次元情報(入力画像)から基本特徴量を抽出し、その後、それらの特徴量を組み合わせ低次元情報に圧縮する。その低次元情報を Decoder 区間で高次元情報、すなわち、入力画像サイズに戻し、意図した情報をピクセル単位で再現する。

本稿では semantic segmentation のピクセル単位の再現精度が低次元特徴量と Decoder の再現精度に起因すると考え、本章で特徴量マップの把握方法を整理する。Decoder 区間の高次元空間への再現プロセスも重要であるが、本論文ではこの部分の構造には詳しく言及せず、特性として考慮する。

3.1 着目する特徴量空間

再現したい情報精度は、最も低次元化した特徴量マップの空間特性に左右されると考える。この低次元特徴量マップに必要な情報が配置されていれば、そこから upsampling された情報も適切な情報である可能性が高くなる。

今回、最も低次元化された特徴量マップとして upsampling 前の pooling 層に着目する。これを更に圧縮し、人間の目で空間配置を確認できるように 2 次元化する。2 次元化する手法としては、PCA, SVD, t-SNE, UMAP 等が代表的であるが、各種法の類似性等も考慮し、本稿では PCA と t-SNE を扱う。

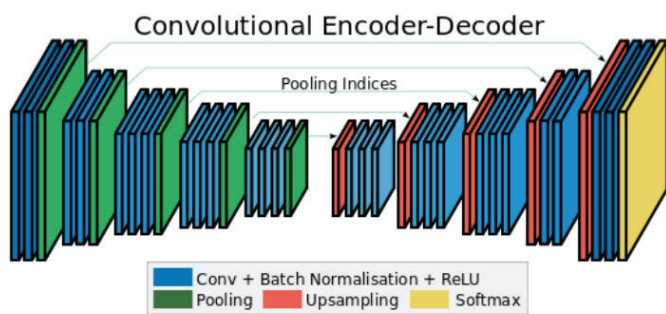


図 3 SegNet のネットワーク構造

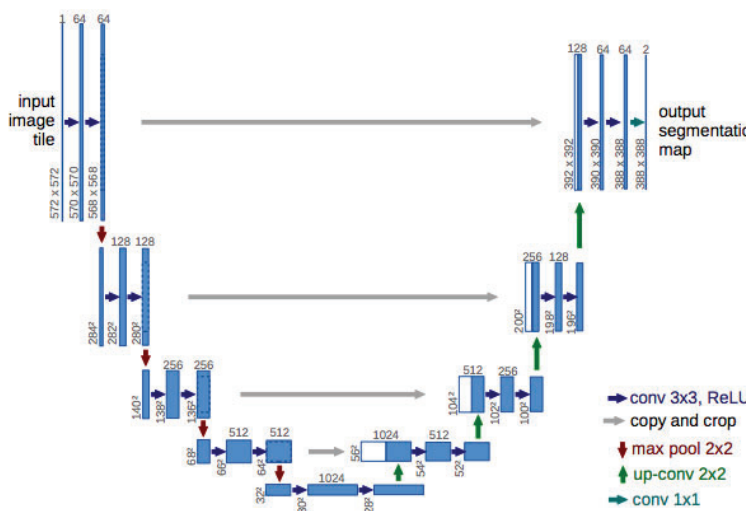


図 4 U-Net のネットワーク構造

3.2 情報圧縮方法の概要

多次元情報を次元圧縮する PCA と t-SNE の概要を記す。

(1) PCA[Hotelling 1933]の概要

多次元空間の特徴量を線形近似経路で次元圧縮していく手法である。データ全体を見て表現力の高い方向に新評価軸を設定していくため、データの全体構造を踏まえた次元圧縮となる。一方で、非線形変換でないため、局所構造を読み取れない等の課題もある。このため、classification や detection、そして、semantic segmentation でもある程度の領域をマーキングする、物体構造の特徴を把握する必要のあるタスクの場合には適切な次元圧縮が可能になると想定される。一方、本稿のひび割れ検出では適切な次元圧縮が出来るかは疑問である。

(2) t-SNE[Maaten 2008]の概要

多次元空間を歪めて複雑なデータの非線形関係を読み解くことも可能である。全体構造よりも局所構造を読み取ることを得意とするが、perplexity の極大化により全体構造を読み取る試みもある。本稿では推奨値である perplexity 5~50 で検討を進め、2 次元化マップを作成する。t-SNE は PCA よりも複雑な局所の特徴量を上手く捉えることができるため、本稿のタスク特性からも解析結果の解釈としては t-SNE をより信用して整理する。

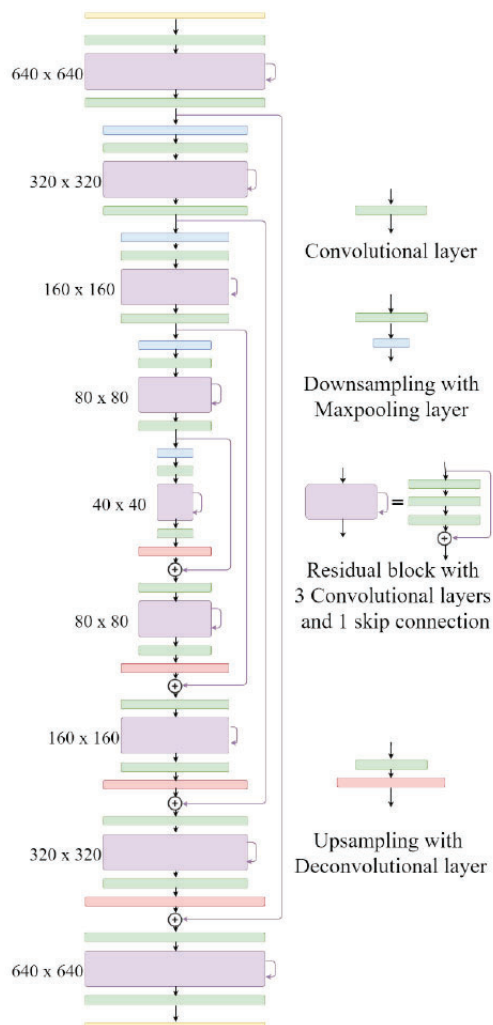


図 5 FusionNet のネットワーク構造

表 1 PCA による特徴量空間の 2 次元化

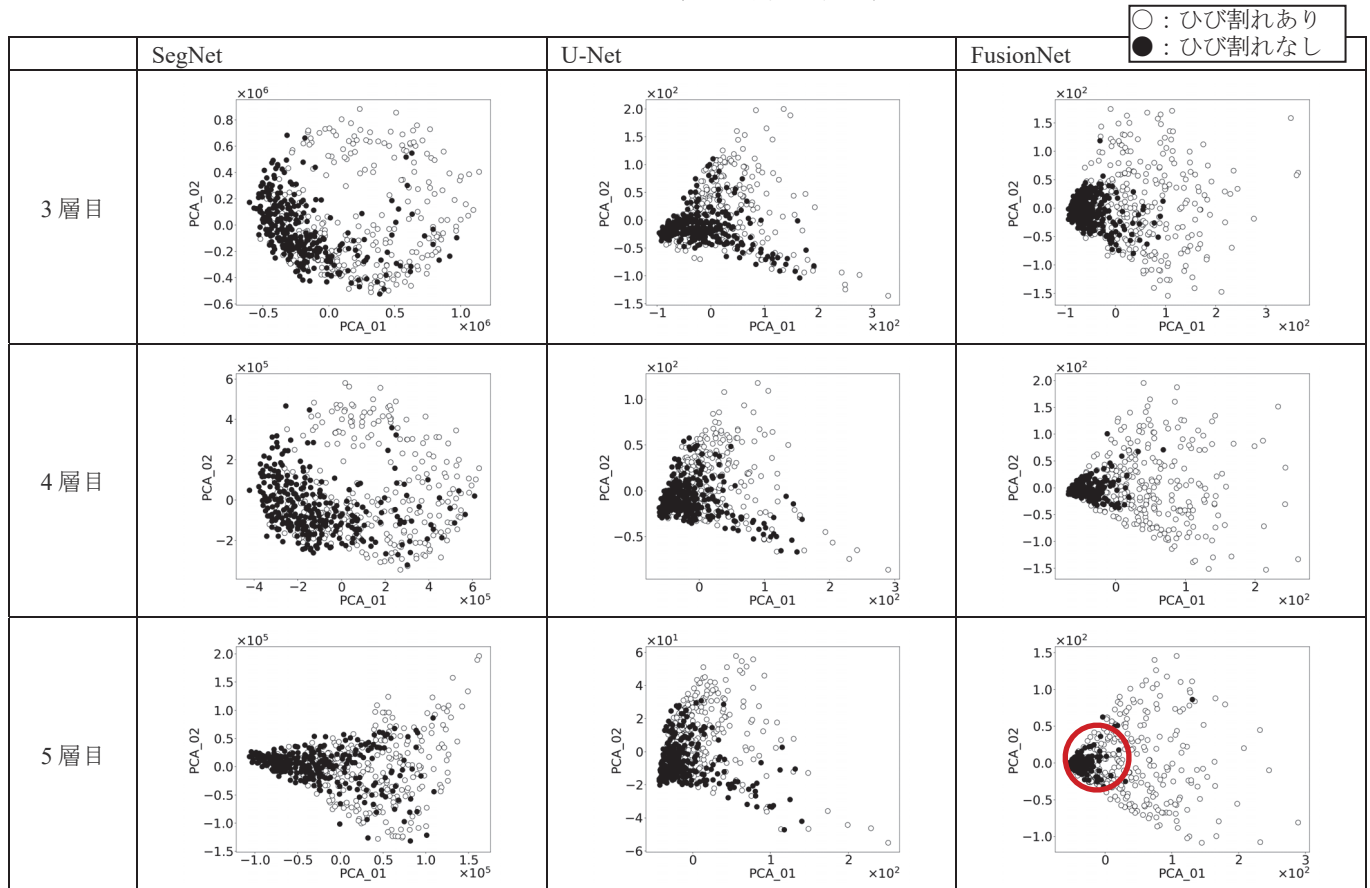


表 2 t-SNE による特徴量空間の 2 次元化

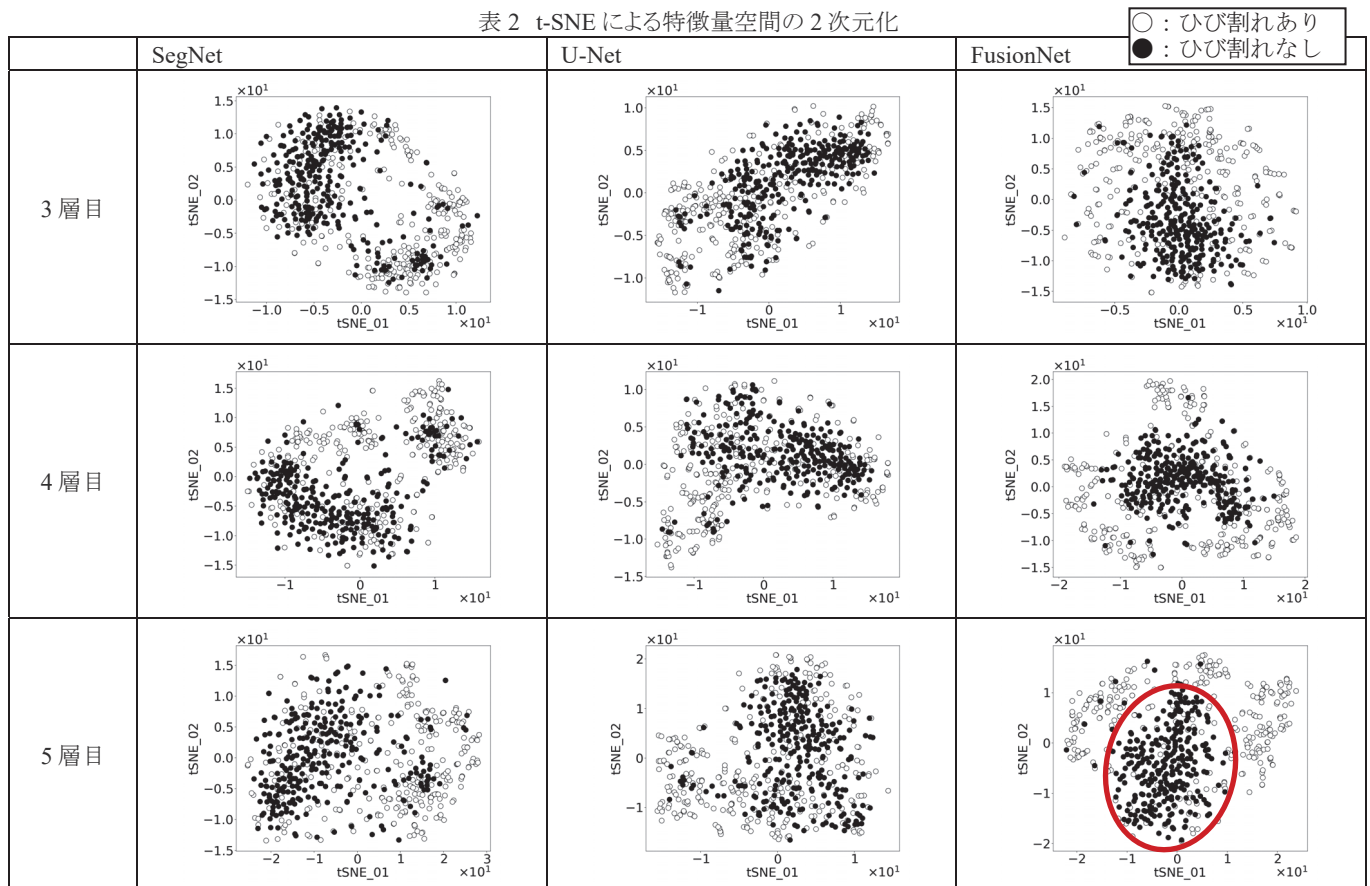


表3 各ネットワークの計算条件一覧

計算条件	条件内容
input_size	(224,224,3)
epoch	50
optimizer	SGD(lr=0.01, decay=1e-4, momentum=0.9)
loss	cross entropy
batch size	16

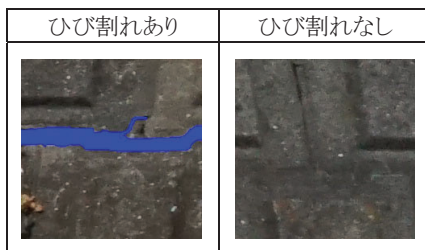


図6 クロップした画像

4. ネットワークの特徴量空間の違い

SegNet, U-Net, FusionNet の低次元化した特徴量マップを更に PCA と t-SNE によって低次元化したものを表 1 と表 2 に作成した(3 層目~5 層目を提示). この特徴量マップは, 河川のコンクリート護岸写真にひび割れのアノテーションを施し, 224×224 にクロップ(図 6 参照)して約 10,000 枚学習し, 600 枚(ひび割れあり 300 枚, ひび割れ無し 300 枚)のテスト画像の計算結果をプロットしたものである. これらの計算条件は表 3 の通りである. 全て同一条件にて 3 つのネットワークを計算している.

この結果を踏まえ, ネットワークの違いによる特徴量マップの違いに言及し, 社会資本施設を対象とする観点からひび割れ検出の Semantic Segmentation 適用時の留意点を示す.

4.1 線形性と非線形性

PCA と t-SNE の解析特性から多次元関係が線形でない限り 2 手法の特徴量マップには違いが出る. ひび割れあり/なしの単純画像とは言え, 例えば, 表 1 と表 2 の FusionNet の 5 層目の PCA と t-SNE の特徴量マップには明確な違いがあり, 多次元特徴量は非線形関係であると分かる. このため, t-SNE との相性が良いというよりは PCA では特徴量の低次元化に限界があることになる. 明確にひび割れあり/なしが区分されているのは FusionNet の t-SNE 結果であり, 大規模構造物の画像解析時に必須となるクロップ画像の局所的微量抽出の場面では, 非線形性を表現できる深層ネットワーク構造が不可欠ということになる.

4.2 局所構造と全体構造

表 1 の PCA の FusionNet5 層目の特徴量マップを見ると, ひび割れなしの特徴量が局所に集まり, 画像面積的には大部分を占めるにも関わらず多様性がない. 一方, ひび割れありの特徴量は広がりを見せ, 局所の多様性が強調されている. これは, クロップ画像学習で河川護岸全体の特徴を掴んでいないためである. 表 2 の t-SNE の FusionNet5 層目の特徴量マップでは, ひび割れなし分布を取り囲むようにひび割れあり分布が広がり, 局所的な特徴量の違いを上手く捉えられている. 一方, SegNet と U-Net の 5 層目の特徴量マップではひび割れありとなしの特徴量が適切に分類されず, その結果が図 6 にも反映されている. ひび割れのような局所構造を捉える情報検出には, Semantic Segmentation の手法は適していると考えられる.

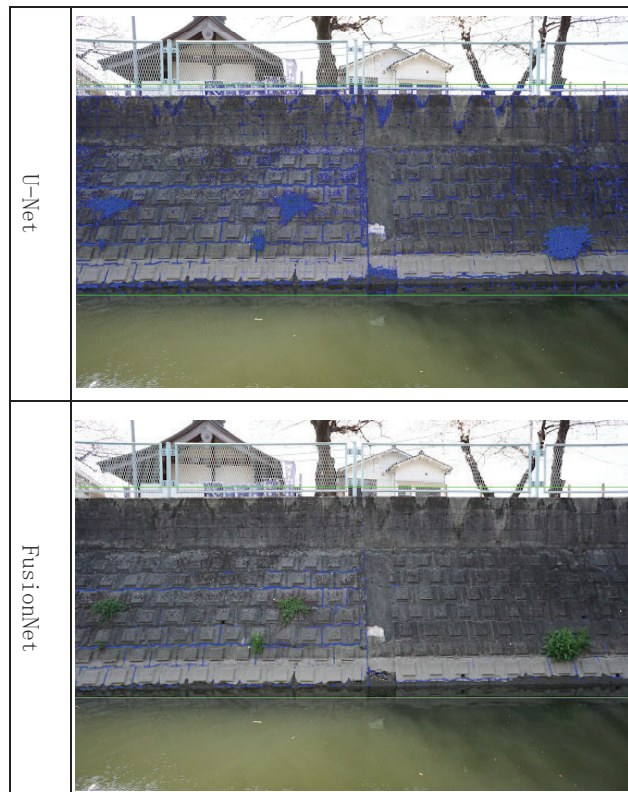


図6 U-Net と FusionNet の出力比較

5. おわりに

各ネットワークの出力画像と特徴量マップの対応を見ると, 出力画像の精度の良さは特徴量マップの分かりやすさと対応していると思われる. 深層ネットワークや教師データの工夫によって特徴量マップをある程度コントロールすることで検出精度を高めることができると考えられる.

一方で, 社会資本施設は構造物であり, 構造物全体の特性を捉えることで検出精度が向上すると考えられる. しかし, 本稿でのクロップ学習や semantic segmentation では局所構造しか捉えられなさそうである. 全体構造も学習できるであろう attention 等の技術についても検討していく必要がある.

参考文献

- [Badrinarayanan 17] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, :SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE transactions on pattern analysis and machine intelligence39(12), 2017.
- [Ronneberger 15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox : U-Net: Convolutional Networks for Biomedical Image Segmentation, in N.Navab, J.Hornegger, W.M.Wells, A.F.Frange(Eds.), Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, 2015.
- [Quan 17] Tran Minh Quan, David G.C.Hildebrand and Won-Ki Jeong :FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics, 2017.
- [Hotelling 1933] Harold Hotelling: Analysis of a Complex of Statistical Variables into Principal Components, Journal of Educational Psychology(24), 1933.
- [Maaten 2008]Laurens van der Maaten, Geoffrey Hinton: Visualizaing Data using t-SNE, Journal of Machine Learning Research 9(2008), 2008.