H-048

クラス情報による誘導が可能な Discrete Flow Matching による データ生成モデルの検討

A Study on Data Generation Model Using Discrete Flow Matching with Class-Conditioned Guidance

藤岡 雅大† 鈴木 海友† 松田 一朗†
Masahiro FUJIOKA Kaiyu SUZUKI Ichiro MATSUDA

1. はじめに

Discrete Flow Matching[1] (DFM) は，連続空間において常微分方程式により生成過程を定義する Flow Matching[2] (FM) を離散空間へと拡張した手法である．DFM は，連続時間マルコフ過程における確率経路の回帰を通じて離散データを生成する，非自己回帰型生成モデルの枠組みであり，言語生成や画像生成といった高次元離散データの評価タスクにおいて高い性能を示している．

生成モデルにおいて出力を制御することは，生成モデルの応用範囲を広げるだけでなく，ユーザの意図に沿った出力を得るために不可欠である．しかし，現行の DFM では，言語翻訳のような状態初期値と出力の対応付けによる条件付け手法は存在するが，クラス情報などの状態以外の入力によって生成結果を誘導する手法が未確立である．

一方，Discrete Diffusion Models[3] (DDM) は，離散時間のマルコフ過程を通じてノイズを付与した離散データの逆過程を学習する手法であり，この手法の出力制御方法として，クラス情報を用いる Classifier-Free Guidance[4] (CFDDM) が提案されている．

本研究では，Classifier-Free Guidance を DFM に適用し，DFM を用いたクラス情報による出力誘導の定式化について検討する．また，提案手法の有効性を確認するために画像生成実験を実施し，生成画像に含まれるクラス情報の妥当性を Fréchet Inception Distance[5]（FID）およびクラス分類モデルを用いたダイバージェンスによって評価した．

2. Discrete Flow Matching[1]

Discrete Flow Matching (DFM) は，連続空間の Flow Matching (FM) の枠組みを離散データ生成に適応した手法であり，離散空間上での非自己回帰型生成モデルフレームワークの一種である．

ここでは，離散空間 $\mathcal{D}$ を K 通りの離散値を持つ d 次元のデータ空間 $\{K\}^d$ とし，$X_t \in \mathcal{D}$ を時刻 $t \in [0,1]$ における確率変数とする．DFM は，ソース分布 p から得られるソースサンプル $X_0 \sim p$ から，ターゲット分布 q に従うターゲットサンプル $X_1 \sim q$ を生成することを目的とする．そのために，p と q を結ぶ確率経路 p_t を定義し，その経路に沿った確率質量関数の時間発展を連続時間マルコフ過程 (CTMC) の形式でモデル化する．具体的には，時刻 t での確率変数 X_t が与えられたとき，時刻 $t+h$ の確率変数 X_{t+h} は 0 から d までの各トークン次元 $i \in \{d\}$ ごとに独立に次のようにサンプリングされる．

$$X_{t+h}^{(i)} \sim \delta_{X_t^{(i)}}(\cdot) + h u_t^{(i)}(\cdot, X_t)$$

ここで，$u_t^{(i)}(\cdot, X_t)$ は時刻 t における状態 X_t が与えられたときの次元 i におけるベクトル場であり，$\delta_x(\cdot)$ は Dirac のデルタ関数である．

2.1 DFM の学習

DFM の学習は，各時刻 t における $u_t^{(i)}(\cdot, X_t)$ を回帰することで行われる．機械学習モデルの損失関数 $\mathcal{L}$ は，CTMC により決定される $u_t^{(i)}(\cdot, X_t)$ と，X_t が与えられたときのモデル出力 $\hat{u_t}^{(i)}(\cdot, X_t; \theta)$ との Bregman ダイバージェンス $D(\cdot,\cdot)$ の，次元 i ごとの和の期待値で表される [6].

$$\mathcal{L}(\theta) = \mathbb{E}_{t,X_t}\left[\sum_{i\in\{d\}} D_{X_t}\left(u_t^{(i)}(\cdot, X_t), \hat{u_t}^{(i)}(\cdot, X_t; \theta)\right)\right] \quad (1)$$

また，任意の状態空間 $\mathcal{Z}$ から得られる確率変数 Z で条件付けた条件付きベクトル場 $u_t^{(i)}(\cdot, X_t \mid Z)$ とモデル出力との Bregman ダイバージェンスで定義される損失関数と，式 (1) の損失関数は，モデルパラメータ θ の勾配が一致することが知られている [6].

$$\nabla_\theta \mathcal{L}(\theta) = \nabla_\theta \mathbb{E}_{t,X_t,Z}\left[\sum_{i\in\{d\}} D_{X_t}\left(u_t^{(i)}(\cdot, X_t \mid Z), \hat{u_t}^{(i)}(\cdot, X_t; \theta)\right)\right]$$

そのため，実装上では Z をソースデータ x_0 とターゲットデータ x_1 とした条件付きベクトル場 $u_t^{(i)}(\cdot, X_t \mid x_0, x_1)$ を用いて損失関数を定義する．

3. Classifier-Free Discrete Diffusion Models[4]

Discrete Diffusion Models (DDM)[3] は，Diffusion Models (DM)[7] を離散データに適用したモデルフレームワークである．DFM と異なり，離散時間のマルコフ過程によるノイズデータ汚染のデノイジングを学習することでデータ生成を行う．

3.1 DDM における Classifier-Free Guidance

DDM では，クラス条件制御を行う手法として Classifier-Free Guidance[4] が提案されている．$x_t \in \mathcal{D} = \{K\}^d$ を離散時刻 $t \in \{T\}$ における状態変数とし，$c \in C$ を状態空間 C 内の条件変数とする．DDM での Classifier-Free Guidance では，状態変数 x_{t+1} の事後分布を，条件付き事後分布 $p_{t+1|C,t}(x_{t+1} \mid c, x_t)$ と非条件付き事後分布 $p_{t+1|t}(x_{t+1} \mid x_t)$ を用いて，各次元 $i \in \{d\}$ ごと独立に定義する．

$$p_{\gamma,t+1|C,t}(x_{t+1} \mid c, x_t) = \prod_{i=1}^{d} \frac{p_{t+1|C,t}(x_{t+1}^{(i)} \mid c, x_t)^\gamma p_{t+1|t}(x_{t+1}^{(i)} \mid x_t)^{1-\gamma}}{\sum_{x'_{t+1}} p_{t+1|C,t}(x'_{t+1} \mid c, x_t)^\gamma p_{t+1|t}(x'_{t+1} \mid x_t)^{1-\gamma}}$$

† 東京理科大学 創域理工学部
Faculty of Science and Technology, Tokyo University of Science

なお，γ は x_{t+1} の c に対する尤度を調節する温度パラメータであり，出力分布の条件忠実性と多様性のトレードオフを調節することができる．

ここで，マスクされた条件変数を入力としたモデル出力を非条件付き生成とみなすことで，条件付き生成と非条件付き生成において同一のモデルパラメータ θ を用いることができる．なお，一般的に学習時には，条件変数をバッチ単位で一定の確率でマスクする．

また，DDM では離散時間間隔を 0 に近づける極限において，データ汚染過程のマルコフ過程が CTMC に収束することが知られている．この点で，DFM と DDM は共通の枠組みに属し，DDM で提案された Classifier-Free Guidance の DFM への適用が期待できる．

4. DFM における出力の条件付け

提案手法では，任意の状態空間 $\mathcal{C}$ 内の条件変数 $c \in \mathcal{C}$ で DFM の条件付けを行う．具体的には，時刻 $t \in [0,1]$ で c と状態変数 $x \in \mathcal{D}$ が与えられたとき，条件付き確率経路 $p_{t+h|C,t}(y \mid x, c)$ を時刻 $t+h$ で状態変数 $y \in \mathcal{D}$ に遷移する条件付き確率経路とし，ベクトル場 $u_{t|C}(y, x \mid c)$ を時刻 $t+h$ で x から y への遷移を表す条件付きベクトル場とする．このとき，条件付けられた DFM (CDFM) における $p_{t+h|C,t}(y \mid x, c)$ は $u_{t|C}(y, x \mid c)$ を用いて次式で表される．

$$p_{t+h|C,t}(y \mid c, x) = \delta_x(y) + h u_{t|C}(y, x \mid c) \tag{2}$$

また，DFM と同様に各トークン次元 $i \in \{d\}$ ごとに独立にサンプリングされると仮定する．

$$\begin{aligned} X_{t+h}^{(i)} &\sim p_{t+h|C,t}^{(i)}(y \mid c, x) \\ &= \delta_{X_t^{(i)}}(\cdot) + h u_{t|C}^{(i)}(\cdot, X_t \mid c) \end{aligned}$$

また，損失関数は c を用いて以下のように書き換えられる．

$$\mathcal{L}(\theta) = \mathbb{E}_{t, X_t, Z, c}\left[\sum_{i \in \{d\}} D_{X_t}\left(u_{t|C}^{(i)}(\cdot, X_t \mid Z, c), \hat{u}_{t|C}^{(i)}(\cdot, X_t \mid c; \theta)\right)\right]$$

5. CDFM における Classifier-Free Guidance

提案手法として，DDM での Classifier-Free Guidance を DFM に適用し，CDFM によるモデル出力が条件変数に対する尤度を高くするように誘導する手法を提案する．

5.1 CDFM における Classifier-Free Guidance の導出

, 式 (2) の条件付き確率経路 $p_{t+h|C,t}(y \mid c, x)$ はベイズの定理を用いて次のように表される．

$$p_{t+h|C,t}(y \mid c, x) = \frac{p_{C|t+h,t}(c \mid y, x)\, p_{t+h|t}(y \mid x)}{p_{C|t}(c \mid x)} \tag{3}$$

式 (3) の変形を踏まえ，温度パラメータ $\gamma \in \mathbb{R}$ を用いて誘導補正した条件付き補正確率経路 $p_{\gamma,t+h|C,t}$ を以下のように定義する．

$$p_{\gamma,t+h|C,t}(y \mid c, x) = \frac{\frac{p_{C|t+h,t}(c|y,x)^{\gamma} p_{t+h|t}(y|x)}{p_{C|t}(c|x)}}{\sum_{y'} \frac{p_{C|t+h,t}(c|y',x)^{\gamma} p_{t+h|t}(y'|x)}{p_{C|t}(c|x)}} \tag{4}$$

式 (4) は，ベイズの定理を用いて次のように変形できる．

$$p_{\gamma,t+h|C,t}(y \mid c, x) = \frac{p_{t+h|C,t}(y \mid c, x)^{\gamma} p_{t+h|t}(y \mid x)^{1-\gamma}}{\sum_{y'} p_{t+h|C,t}(y' \mid c, x)^{\gamma} p_{t+h|t}(y' \mid x)^{1-\gamma}} \tag{5}$$

式 (5) より，$p_{\gamma,t+h|C,t}(y \mid c, x)$ は，DDM における Classifier-Free Guidance と同様に，CDFM の条件付き確率経路 $p_{t+h|C,t}(y \mid c, x)$ と DFM の非条件付き確率経路 $p_{t+h|t}(y \mid x)$ を組み合わせた形となる．また，γ は DDM における Classifier-Free Guidance と同様に y における c の尤度を調節するパラメータとして作用する．

また，$p_{t+h|C,t}(y \mid c, x)$ と $p_{t+h|t}(y \mid x)$ はそれぞれ y のトークン次元 $i \in \{d\}$ ごとに独立であるため，式 (5) は各トークン次元 i ごとに分解できる．

$$p_{\gamma,t+h|C,t}(y^{(i)} \mid c, x) = \frac{p_{t+h|C,t}(y^{(i)} \mid c, x)^{\gamma} p_{t+h|t}(y^{(i)} \mid x)^{1-\gamma}}{\sum_{y^{(i)}} p_{t+h|C,t}(y^{(i)} \mid c, x)^{\gamma} p_{t+h|t}(y^{(i)} \mid x)^{1-\gamma}}$$

5.2 サンプリング方法

サンプリング時には，学習済み CDFM と DFM を用いて十分大きい時間ステップ数 N のオイラー近似によりサンプリングする．各サンプリング時刻 $t \in \{1/N, 2/N, ..., 1\}$ において，CDFM と DFM には同一の状態変数と条件変数を入力し，それぞれのモデルから得られた出力から条件付き補正確率経路を算出する．

$$X_{t+h} \sim p_{\gamma,t+h|C,t}(\cdot \mid c, X_t)$$

また，DDM での Classifier-Free Guidance と同様に，マスクされた条件変数の入力に対応する出力を DFM の出力とすることで，DFM と CDFM は同一のモデルを使用できる．

6. 実験

提案手法のクラス情報による出力誘導効果を検証するため，時刻ステップ数を 1028 とし，温度パラメータ γ を {0.0, 0.2, 0.5, 0.8, 1.0, 1.2, 1.5, 2.0, 3.0} の 9 通りで画像生成実験を行った．実験では，離散値を持つデータセットとして，図 1 のように画素値を固定閾値 0.5 で 0 と 1 に 2 値化した MNIST データセット (Binarized MNIST) を用いた．また，CDFM と DFM 共通の機械学習モデルとして U-Net[8](パラメータ数約 113M) を実装した．モデルの学習には Binarized MNIST の学習データセットを用い，クラス情報としては画像ラベルと，画像ラベルとは異なる特定のカテゴリを割り当てたマスクラベルを 7:3 の割合で入力して，100 エポックの学習を行った．また，評価時には，Binarized MNIST の評価ラベルデータを用い，CDFM と DFM のクラス情報としてそれぞれ画像ラベルとマスクラベルを入力し，画像生成を行った．

6.1 生成画像の FID 評価

6.1.1 実験方法

γ の各値に対して，生成画像群と評価データセットとの間で FID[5] (Fréchet Inception Distance) による評価を行った．FID は，クラスラベル別に計算したマクロ平均と，データセット全体の画像群に基づくマイクロ平均の 2 種類を算出した．なお，FID は生成画像群と実画像群の分布の Fréchet 距離を測る指標であり，値が小さいほど分布の類似度が高いことを示す．

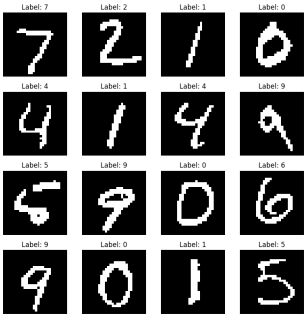

図 1: 閾値 0.5 の Binarized MNIST データセット

6.1.2 実験結果

図 2 に γ の各値に対するクラスごとの FID を，図 3 にクラスごとの FID のマイクロ平均とデータセット全体のマクロ平均を示す．

図 2 から，全てのクラスにおいて $\gamma = 1.2$ 付近で FID が最小を取る凸状の傾向が見られた．

また，図 3 から，$0 \leq \gamma \leq 0.5$ の範囲では，γ が小さいほど FID のマクロ平均は大きくなる一方，FID のマイクロ平均はマクロ平均ほどの増加は見られなかった．これは，クラス条件付けのない DFM では全クラスの分布を再現するように学習されるため，γ によらず FID のマイクロ平均は大幅な変化は見られない一方，γ が 0 に近いほどクラス誘導の影響が小さくなり，生成画像と評価データセットのクラス分布が乖離していることを示唆している．また，$1.2 \leq \gamma \leq 3.0$ の範囲では，γ の増加に伴い FID のマクロ平均とマイクロ平均ともに増加する傾向が見られた．これは，γ が大きくなるほどクラス情報の影響が強くなり，生成画像の分布がクラスに対して尤もらしい分布に偏ったことを示唆していると考えられる．

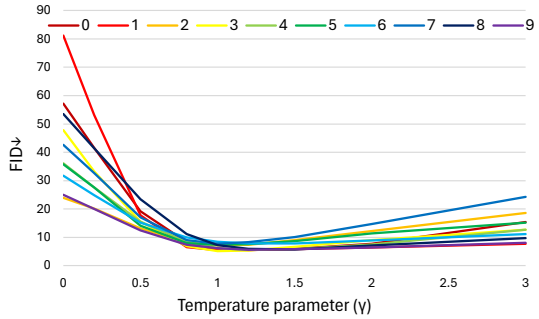

図 2: クラスラベル 0〜9 ごとの FID

6.2 クラス分類結果の KL ダイバージェンス評価

6.2.1 実験方法

6.1 章で生成した画像を事前学習したクラス分類モデルに入力し，出力されたクラス尤度と正解ラベルの one-hot ベクトル間の KL ダイバージェンスを，γ ごとに算出した．クラス分類モデルには WideResNet-28-10[9] を用い，Binarized MNIST の学習データセットを用いて 100 エポックの学習を行った．また，クラス分類モデルの評価性能として，Binarized MNIST 評価データセットに対する分類精度 (%) および正解ラベルとの KL ダイバージェンスを表 1 に示す．

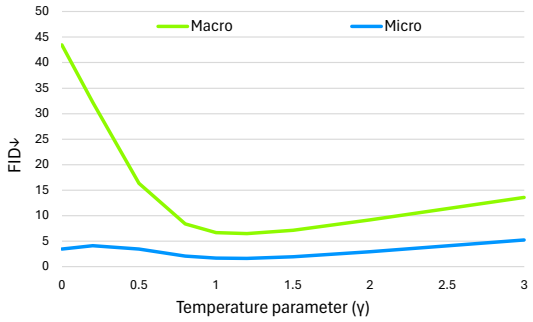

図 3: クラスごとの FID のマイクロ平均とマクロ平均

表 1: ラベルごとの分類精度と KL ダイバージェンス

Label	Accuracy (%)	KL Divergence
0	99.90	0.23×10^{-2}
1	99.82	0.51×10^{-2}
2	99.61	0.60×10^{-2}
3	99.90	0.41×10^{-2}
4	99.49	2.56×10^{-2}
5	99.33	4.15×10^{-2}
6	99.16	3.23×10^{-2}
7	99.12	3.51×10^{-2}
8	99.69	0.76×10^{-2}
9	98.81	4.79×10^{-2}
Average	99.48	2.08×10^{-2}

6.2.2 実験結果

生成画像に対するクラス尤度と正解ラベルとの KL ダイバージェンスのマクロ平均，および表 1 に示す評価データセットに対するクラス尤度との KL ダイバージェンスのマクロ平均を，γ ごとにプロットした結果を図 4 に示す．図 4 より，$\gamma = 0$ から $\gamma = 3.0$ にかけて KL ダイバージェンスのマクロ平均が単調に減少する様子が確認できた．これは，γ の増加によりクラス情報の寄与が強まり，生成画像のラベルに対する尤度が高まることを示唆していると考えられる．また，$\gamma \geq 1.5$ 付近からは，生成画像のクラス分類による KL ダイバージェンスが，評価データセットのクラス分類による値を下回ることが確認できた．

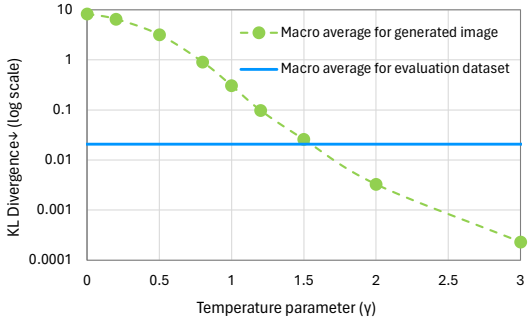

図 4: 生成画像と評価データセットの KL ダイバージェンス

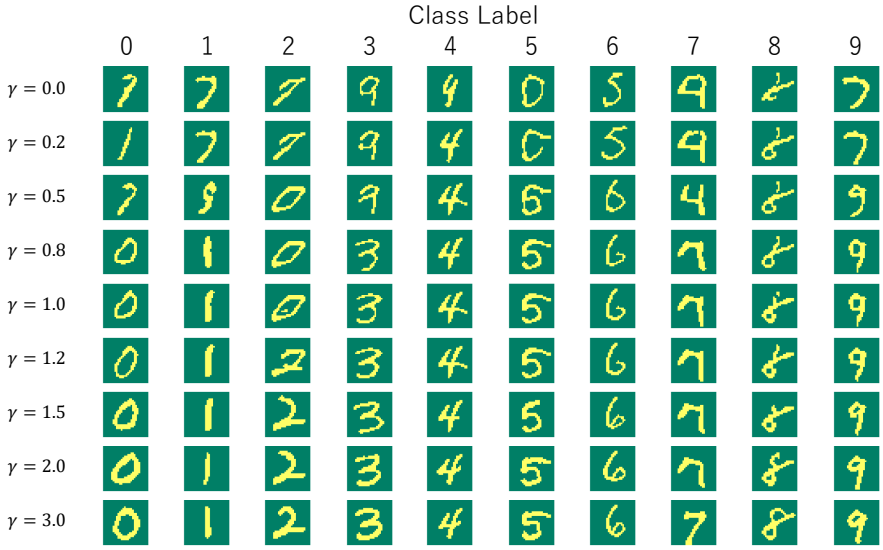

図 5: γ ごとの画像生成結果

6.3 定性評価

6.3.1 実験方法

γ ごとに共通の初期状態とクラス情報を与えて画像生成を行い，生成画像を定性的に比較した．なお，付与する初期状態とクラス情報の組み合わせは，0 から 9 の各クラスラベルごとの 10 通りで比較した．

6.3.2 実験結果

図 5 に γ ごとの生成画像を示す．DFM による非条件付き生成である $\gamma = 0.0$ のときは，クラス情報に依存しない画像が生成され，複数の画像において複数のラベルの特徴が混在する様子が確認できた．$\gamma = 0.0$ から 1.0 にかけては，生成画像が次第にクラス情報に対応する数字の形状へと近づいていく様子が確認できた．一方，CDFM による条件付き生成である $\gamma = 1.0$ において，ラベル 8 やラベル 2 の生成画像ように，数字の形状が不明瞭であったりクラス情報と異なる数字が生成される例が確認できた．しかし，$\gamma = 1.0$ から 3.0 にかけては，全てのラベルの出力画像において，数字の形状がクラス情報と整合的な形状に近づいていく様子が確認できた．

7. おわりに

本研究では，Discrete Flow Matching にクラス情報を付与し，サンプリング時において，クラス情報が出力へ与える影響を温度パラメータ γ を用いて調整する方法を定式化した．評価データセットと生成画像の FID の比較より，γ の値により出力画像の分布を誘導できることを確認した．また，生成画像の分類器モデル出力とラベルとの KL ダイバージェンスの比較と生成画像の定性評価より，生成画像のラベルに対する尤もらしさを制御できるとも明らかにした．

今後の展望として，言語ドメインを始めとした画像以外の離散データでの評価や，より自由度の高い条件付けの手法検討，Guided Discrete Diffusion Model との性能，特性比較調査などが挙げられる．

参考文献

[1] I. Gat, T. Remez, N. Shaul, F. Kreuk, R. T. Q. Chen, G. Synnaeve, Y. Adi, and Y. Lipman, “Discrete Flow Matching,” Proc. 38th Conf. on Neural Information Processing Systems (NeurIPS 2024), Dec. 2024.

[2] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, “Flow Matching for Generative Modeling,” Proc. 11th Int. Conf. on Learning Representations (ICLR 2023), Apr. 2023.

[3] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, and R. van den Berg, “Structured Denoising Diffusion Models in Discrete State-Spaces,” Proc. 35th Conf. on Neural Information Processing Systems (NeurIPS 2021), pp. 17981–17993, Dec. 2021.

[4] Y. Schiff, S. S. Sahoo, H. Phung, G. Wang, S. Boshar, H. Dalla-torre, B. P. de Almeida, A. Rush, T. Pierrot, and V. Kuleshov, “Simple Guidance Mechanisms for Discrete Diffusion Models,” Proc. Causal and Object-Centric Representations for Robotics (CORR Workshop at CVPR 2024), vol. abs/2401.00001, Jun. 2024.

[5] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” Proc. 31st Int. Conf. on Neural Information Processing Systems (NeurIPS 2017), pp. 6626–6637, Dec. 2017.

[6] Y. Lipman, M. Havasi, P. Holderrieth, N. Shaul, M. Le, B. Karrer, R. T. Q. Chen, D. Lopez-Paz, H. Ben-Hamu, and I. Gat, “Flow Matching Guide and Code,” arXiv preprint arXiv:2412.06264, Dec. 2024.

[7] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep Unsupervised Learning using Nonequilibrium Thermodynamics,” Proc. 32nd Int. Conf. on Machine Learning (ICML 2015), pp. 2256–2265, Jul. 2015.

[8] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Lecture Notes in Computer Science, vol. 9351, Proc. 18th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), pp. 234–241, Sept. 2015.

[9] S. Zagoruyko and N. Komodakis, “Wide Residual Networks,” Proc. British Machine Vision Conf. (BMVC 2016), Sept. 2016.