

## 下降型減衰周期的学習率 Scheduler を用いた

## 生成画像を含む混合データセット学習の品質低下防止

## A method for preventing quality degradation when training mixed datasets containing generated images using Descending Damped Cyclical Learning Rates Scheduler.

杉浦 陸空<sup>†</sup> 中村 和幸<sup>†</sup>  
Riku Sugiura Kazuyuki Nakamura

## 1. 序論

## 1.1 背景・動機づけ

今日では生成 AI 技術の発達により実画像と見分けのつかないほど高精度な生成画像を低コストで迅速に大量に生成することが可能になっている。生成画像は様々な用途で用いられているが、積極的な利活用に伴い機械学習の学習データセットとして実画像の一部を生成画像で代用した混合データセットを用いることでデータ不足を解消する例が散見されるようになった。一方で、生成モデルの学習においてはそのような混合データセットを用いた学習を行う場合に予期せぬ品質低下を招く可能性がある。特に学習率に着目すると、本研究セクション 3 の結果では、実データセット学習時には概ねどの学習率設定でも各データセットにおいて一定程度の学習精度を得ることができ、学習率設定は学習における汎用性と精度向上の大きなトピックではなかったが、実・混合データセット双方の学習が一般化されるケースにおいては、各データセットにおいて学習率設定による結果の良し悪しが顕著に現れることから、学習率設定は大きく生成モデル学習の汎用性と精度の欠落に起因する可能性があることを示している。この結果は実・混合データセット双方の学習の一般化において、今までのモデルが汎用的に用いられなくなること、あるいは学習率チューニングコストが増大する可能性を示唆している。そのため、対策として実・混合データセット双方に対して、低い学習率チューニングコストで学習汎用性を確保し、高精度の学習を実現する学習率 scheduler に需要があると考えられる。

## 1.2 関連研究と位置付け

Alemohammad らの研究[1]では、生成モデルの学習に前世代から生成したデータを次世代の学習に用いた検証を Stylegan2[2]等の生成モデルの学習を例におこなっている。結果として生成データのみデータセット学習時と初期学習に用いた実データと前世代の生成データを混合させたデータセット学習時に FID(Frechet Inception Distance)が悪くなった。また、前世代の生成画像と新しい実データで構成された混合データセット学習時には一定以上の実データを含む混合データセットを用いた場合のみ FID が悪くならなかった。結論として、生成モデルの学習時には学習データになるべく実画像を確保することの必要性を訴えている。

本研究の位置付けとして、学習済み生成モデルを所望のデータセットの転移学習により強化するケースにおいて、学習データセット確保のために高精度な大規模生成モデルである Stable Diffusion[3]を用いて生成した画像を用いて混

合データセットを構成するようなケースを想定し、その場合の危険性と改善策を示すこととする。

また、Scheduler は学習が進むごとに学習率を変化させることで学習精度向上に貢献してきた。Smith らの研究[4]では、学習率の基準下限値と上限値の設定を行い、その 2 つの値の間を周期的に学習率が変化する周期的学習率(Cyclical Learning Rates)を提案し、一部のケースにおける学習効率や精度を大幅に向上させた。Pytorch では周期的学習率を実装した Scheduler である CyclicalLR が提供されている。

本研究では CyclicalLR を応用することで下降型減衰周期的学習率 Scheduler を実装し、生成画像を含む混合データセット学習における有用性を示す。

## 1.3 目的

本研究では、いくつかの学習率設定において、複数の混合データセットで stylegan2 を転移学習させた時の FID 低下の様子を示し、学習率設定の課題と混合データセット学習の危険性を確認する。また、提案する下降型減衰周期的学習率 Scheduler の導入によりチューニングコストを抑え、実・混合データセット双方の転移学習において他の学習率設定を使用した場合よりも FID において高い学習汎用性、精度を示すことを目指す。

## 1.4 貢献

本研究では、提案する scheduler の導入により実画像の代用で一部生成画像を含むような混合データセットと実データセット双方の学習を汎用的に高精度で行う手法を提供する。生成 AI の積極的な利活用に伴い一般化が進む混合データセット学習の課題に対して、学習率 Scheduler 開発により対策を試みる点において新規性が高く、導入ハードルが低い点で有用性があり、今後更なる生成 AI データの利活用の促進に貢献をもたらす 1 つの手法になりうると思われる。

## 2. 提案手法・評価方法

## 2.1 データについて

実データとしてサイズ 512×512 の CelebAMask-HQ データセットを用いる。また、生成画像として Stable Diffusion[3]により生成したサイズ 512×512 の顔画像を用いる。生成画像に関して、実画像に似ていない全身画像や 2 人以上写っている画像、絵画に似た画像等を取り除き、実画像と区別がつかない程度の高精度画像を使用した。

## 2.2 使用するネットワークについて

本研究では NVIDIA の NVIDIA Research Project が公開している FFHQ データセット学習済み stylegan2[2]を用いた転移学習を行う。オプティマイザは Adam を使用している。

<sup>†</sup> 明治大学 Meiji University

### 2.3 下降型減衰周期的学習率 Scheduler

本研究では stylegan2 に図 1 に示すような下降型減衰周期的学習率 scheduler を用いることで混合データセット学習に対する対策を講じる。本研究で用いる手法は pytorch の CyclicLR を応用し、下限値(base\_lr)と上限値(max\_lr)の入力から Scheduler の更新時に以下の式(1)(2)により新たに下降型下限値(newbase\_lr)と減衰周期振幅(D\_cycle)を定義している。

$$\text{newbase\_lr} = \frac{\text{base\_lr}(\text{total\_epoch} - \text{last\_epoch})}{(\text{total\_epoch} + \text{last\_epoch})} \quad (1)$$

$$\text{d\_cycle} = \frac{(\text{max\_lr} - \text{base\_lr})(\text{total\_epoch} - \text{last\_epoch})}{(\text{total\_epoch} + \text{last\_epoch})} \quad (2)$$

ここで、total\_epoch はトータルの学習エポック数、last\_epoch は Scheduler 更新時点の学習エポック数である。

提案手法の特徴としては次の 3 つが挙げられる。

1 つ目に周期性を持っている点である。関連研究で提案されている Cyclical Learning Rate[4]では図 2 のように学習率を周期的に基準下限値と上限値の間で変化させるものがある。いくつかのケースにおいては周期性を持たせることにより高精度な学習を実現できることに加え、学習を高速化できることが示されている。また、周期的変化により学習率のチューニングコストを下げることができる。提案手法でも学習効率、チューニングコストの観点から周期性を取り入れることにした。

2 つ目に減衰性を持っていることである。Cyclical Learning Rate では、学習率の周期振動を減衰させることで学習を安定させ、加えて既存の学習率減少型 Scheduler と同等以上の学習精度結果が示されている。また、上限に関する初期設定依存性を緩和し、チューニングコストを低減させることができるため、本研究でも減衰性を取り入れた。

3 つ目に下限値が下降型の scheduler であることである。既存のいくつかの scheduler では、学習率を序盤に高く、終盤に低くするような設定により、学習最適解への収束を実現している。提案手法においては振幅上限の下降性(減衰性)に加え、振幅下限の下降性を取り入れることによって高い学習効率による学習最適解への収束を実現させるとともに、上限・下限双方の初期設定依存性を緩和させ、チューニングコストを低減させる狙いがある。よって、本研究では学習率の振幅上限値を下降させる減衰性に加え、振幅下限値の下降性により学習効率向上、チューニングコスト低減を意図して下降性を導入した。

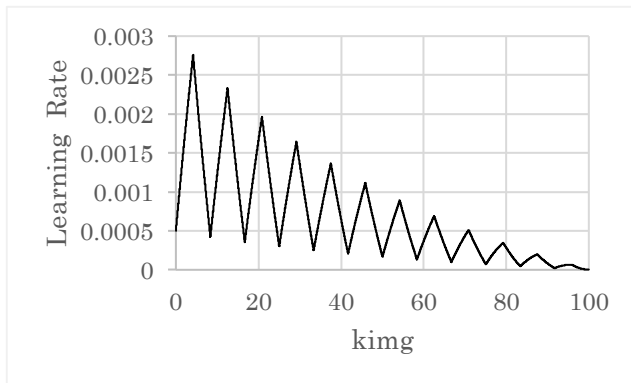


図 1 下降型減衰周期的学習率(DDCLR)の学習率推移

### 2.4 比較対象

以下の表 1 の 8 つの学習率および scheduler 比較する。図 2, 3, 4 は 100king(king:千枚単位)学習時の scheduler の学習率推移の様子である。以下より表の()内の略称を用いる。

DCLR は式(2)を使用することで減衰性を持たせている。また、ExpLR の学習率は以下の式(3)のように更新される。

$$\text{lr} = \text{first\_lr} * \left\{ e^{\frac{\ln(\text{final\_lr}/\text{first\_lr})}{\text{total\_epoch}}} \right\}^{\text{last\_epoch}} \quad (3)$$

ここで、total\_epoch はトータルの学習エポック数、last\_epoch は Scheduler 更新時点の学習エポック数、first\_lr は初期学習率、final\_lr は学習率収束値である。

表 1 比較する学習率・Scheduler と学習率設定

学習率・Scheduler	学習率設定
固定学習率(Fixed)	lr=0.0025
周期的学習率(CLR)	base_lr=0.0005, max_lr=0.0030
下降型減衰周期的学習率(DDCLR)	base_lr=0.0005, max_lr=0.0030
減衰周期的学習率(DCLR)(1)	base_lr=0.0005, max_lr=0.0030
減衰周期的学習率(DCLR)(2)	base_lr=0.0001, max_lr=0.0030
減衰周期的学習率(DCLR)(3)	base_lr=0.0000, max_lr=0.0030
指数関数型学習率(ExpLR)(1)	first_lr=0.0030, final_lr=0.0005
指数関数型学習率(ExpLR)(2)	first_lr=0.0030, final_lr=0.0001

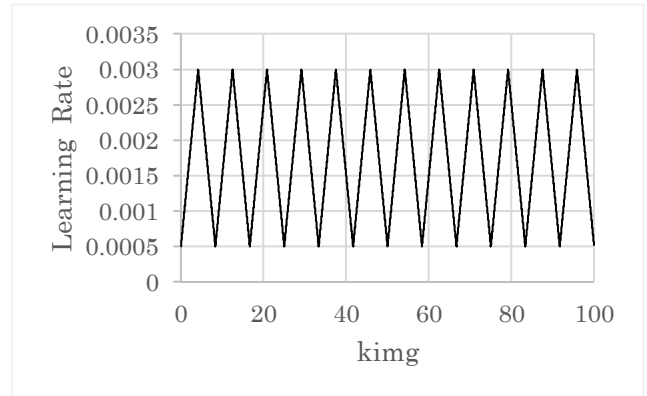


図 2 周期的学習率(CLR)の学習率推移

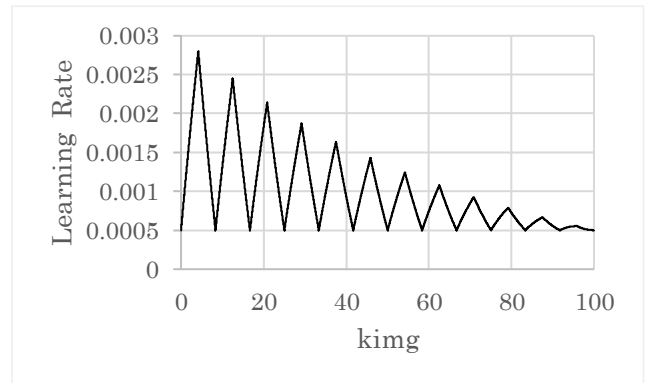


図 3 減衰周期的学習率(DCLR(1))の学習率推移

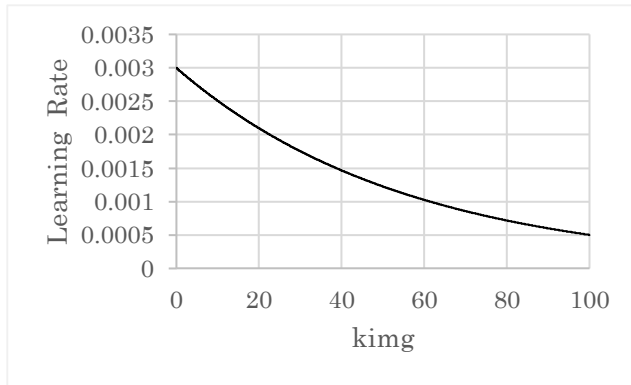


図 4 指数関数型学習率(ExpLR(1))の学習率推移

学習に関して、周期型の Scheduler は学習終了時まで周期を保ちながら 12 周期するように設定している。また、表 1 の学習率設定に関して、設定による結果の偏りを防ぐため、DCLR と ExpLR は複数の学習率設定値を用いて検証を行う。

## 2.5 評価方法

### 2.5.1 前提条件

本研究では stylegan2 を使用し、合計学習枚数が 100king になるまでの学習結果の FID 推移を検証するものとする。また、FID の計算は 50king の生成画像データセットと表 1 に示す固定の実画像データセットを使用する。

### 2.5.2 データセット

以下の表 2 に示すデータセットを検証する。Mix-dataset1~5 は CelebAMask-HQ1~5 のデータセットのうち、200 枚を Stable Diffusion 生成画像で置き換えたものである。

### 2.5.3 実験方法

本研究では表 2 に示す CelebAMask-HQ、Mix-dataset の 5 つずつある学習データセットに対し、FFHQ データセット学習済み stylegan2 を用いて合計枚数 100king で学習を終了する設定の転移学習を番号順に連続的に行い、40, 60, 80, 100king 学習時におけるネットワークの生成画像 50king と表 1 に示すデータセットとの FID を算出する。

## 3. 結果

表 3, 4 に各データセットの 100king 学習時の FID 結果、表 5, 6 に 40king, 60king, 80king, 100king 学習時の最小 FID を示す。表中の数字の太線は各データセットの上位 2 つ、下二重線は各データセットの下位 2 つを表す。

CelebAMask-HQ の検証では表 3, 5 より、DDCLR, DCLR(1), CLR が良い結果を示している。一方で、減衰周期型の DCLR(2)の精度が悪くなっており、DCLR の中で初期値設定による結果のばらつきが見られる。また、Fixed, DCLR(3), ExpLR(1)はデータセットごとに精度の良し悪しにばらつきがあることが示されている。しかしながら、概ね全ての精度が悪くないことがわかる。

Mix-dataset での検証では表 4, 6 より DDCLR の精度が安定して高かった。また、DCLR(1)(2)(3)と ExpLR(1)(2)の学習精度が学習するデータセットによって良し悪しが分かれており、初期設定の難しさが示された結果となった。一方、ほとんどのデータセットにおいて CLR と Fixed の精度が悪

くなっており、特に CLR に関しては、実データセット学習時は有用性が高かったが、混合データセット学習時には他の設定に比べ大きく精度を落とした結果となった。

表 2 学習データセット

	画像数	実データ	生成画像	FID比較対象
CelebAMask-HQ1	1000	1000		0 CelebAMask-HQ1
CelebAMask-HQ2	1000	1000		0 CelebAMask-HQ2
CelebAMask-HQ3	1000	1000		0 CelebAMask-HQ3
CelebAMask-HQ4	1000	1000		0 CelebAMask-HQ4
CelebAMask-HQ5	1000	1000		0 CelebAMask-HQ5
Mix-dataset1	1000	800	200	200 CelebAMask-HQ1
Mix-dataset2	1000	800	200	200 CelebAMask-HQ2
Mix-dataset3	1000	800	200	200 CelebAMask-HQ3
Mix-dataset4	1000	800	200	200 CelebAMask-HQ4
Mix-dataset5	1000	800	200	200 CelebAMask-HQ5

表 3 CelebMask-HQ の 100king 転移学習時の FID

	Dataset1	Dataset2	Dataset3	Dataset4	Dataset5	平均
Fixed	20.420	19.132	<u>20.810</u>	<b>19.675</b>	19.989	<u>20.005</u>
CLR	20.487	<b>18.028</b>	20.235	20.050	19.686	19.697
DDCLR	<b>19.716</b>	18.560	19.944	<b>19.569</b>	<b>19.572</b>	<b>19.472</b>
DCLR(1)	<b>19.529</b>	18.929	<b>19.935</b>	19.862	<u>20.463</u>	19.744
DCLR(2)	<u>20.858</u>	<u>19.391</u>	<u>21.556</u>	19.866	<u>20.950</u>	<u>20.524</u>
DCLR(3)	<u>20.545</u>	<b>18.543</b>	<b>19.790</b>	<u>20.166</u>	<b>19.250</b>	<b>19.659</b>
ExpLR(1)	19.829	18.869	20.318	<u>20.769</u>	20.012	19.959
ExpLR(2)	19.872	<u>19.567</u>	20.684	20.054	19.588	19.953

表 4 Mix-dataset の 100king 転移学習時の FID

dataset	1	2	3	4	5	平均
Fixed	<u>24.134</u>	<u>23.711</u>	<u>27.033</u>	<u>25.152</u>	<u>24.195</u>	<u>24.845</u>
CLR	<u>24.194</u>	<u>23.429</u>	<u>25.288</u>	<u>26.497</u>	<u>26.377</u>	<u>25.157</u>
DDCLR	<b>22.482</b>	<b>20.549</b>	<b>23.108</b>	<b>22.003</b>	<b>21.552</b>	<b>21.939</b>
DCLR(1)	23.291	21.290	24.685	<b>22.725</b>	22.402	22.879
DCLR(2)	<b>22.334</b>	21.234	24.210	23.203	22.368	22.670
DCLR(3)	23.705	<b>20.724</b>	23.496	23.670	<b>21.483</b>	<b>22.616</b>
ExpLR(1)	23.890	22.251	23.700	24.152	23.486	23.496
ExpLR(2)	23.291	21.526	<b>23.404</b>	22.900	23.085	22.841

表 5 CelebAMask-HQ 転移学習時の最小 FID

dataset	1	2	3	4	5	平均
Fixed	<u>20.139</u>	18.089	<u>19.382</u>	<b>18.271</b>	<u>19.989</u>	<u>19.174</u>
CLR	<b>19.272</b>	<b>17.397</b>	19.100	<b>18.194</b>	18.716	<b>18.536</b>
DDCLR	19.716	<b>17.925</b>	19.132	18.807	18.877	18.891
DCLR(1)	19.529	17.961	<b>18.781</b>	18.759	<b>18.597</b>	<b>18.726</b>
DCLR(2)	19.742	<u>18.388</u>	<u>19.718</u>	<u>19.222</u>	19.113	<u>19.237</u>
DCLR(3)	<u>20.087</u>	17.939	19.050	19.022	<b>18.620</b>	18.944
ExpLR(1)	<b>19.068</b>	18.127	<b>18.830</b>	<u>19.167</u>	<u>19.516</u>	18.942
ExpLR(2)	19.754	<u>18.842</u>	19.121	19.141	18.752	19.122

表 6 Mix-dataset 転移学習時の最小 FID

dataset	1	2	3	4	5	平均
Fixed	22.414	<u>21.865</u>	<u>23.872</u>	<u>24.654</u>	<u>24.195</u>	<u>23.400</u>
CLR	<u>22.467</u>	<u>22.400</u>	<u>24.585</u>	<u>24.128</u>	<u>23.554</u>	<u>23.427</u>
DDCLR	21.540	<b>20.179</b>	<b>22.055</b>	<b>21.347</b>	<b>21.476</b>	<b>21.319</b>
DCLR(1)	<b>21.431</b>	20.917	22.529	22.725	22.389	21.998
DCLR(2)	22.334	21.234	22.351	<b>21.419</b>	21.991	21.866
DCLR(3)	<u>22.731</u>	<b>20.341</b>	22.586	21.429	<b>21.483</b>	<b>21.714</b>
ExpLR(1)	22.188	20.352	<b>22.126</b>	22.714	22.826	22.041
ExpLR(2)	<b>21.430</b>	21.156	23.404	22.564	22.264	22.163

また、Mix-dataset 転移学習時に精度が良かった DDCLR, DCLR(3), CelebAMask-HQ データで良い結果となった DCLR(1), Mix-dataset と CelebAMask-HQ でどちらも精度が悪かった Fixed, CelebAMask-HQ で精度良かったものの Mix-dataset で精度が低くなった CLR の 5 つについての Mix-dataset 学習時の FID 推移を表したものが図 5 である。図 5 より、各学習率設定による FID 推移の良し悪しのばらつきが顕著に示されていることがわかる。Fixed と CLR は常に FID の値が高く、急激な増減を繰り返している。

## 4. 考察

### 4.1 混合データセット学習の危険性について

本研究の結果より、実データセット学習で有効性を示していた一部の学習率設定において、混合データセットの学習時に他の設定と比較し FID を大きく低下させるケースがあることより、実・混合データセット双方の学習の一般化はモデルの学習汎用性と精度を大きく損なう危険性があると考えられる。表 3, 5 より、実データセットを用いた場合の FID はどの学習率設定を用いた場合でもおおよそ良い結果を示し、結果のばらつきは混合データセット検証時と比較し、大きくなかった。しかしながら、表 4, 6 が示す通り、混合データセット学習時は全体的に FID が悪くなっていることに加え、学習率設定によって悪化の仕方が大きく異なり、Fixed と CLR は他の学習率設定よりも大きく精度を損なうような結果が見られた。特に、CLR に着目すると実データセット学習時には他の学習率設定と同等以上の精度を示しているにも関わらず、混合データセット学習時に大きく精度を損なった。このように一部ケースで学習率設定が実データのみの場合に有用性が確認できる場合でも、一部を生成データに置き換えるだけで大きく学習汎用性と精度を損なう可能性がある点で、我々は混合データセット学習に対する十分な危険性を認識する必要があると考える。

### 4.2 下降型減衰周期型学習率の導入利点について

本研究で提案する下降型減衰周期型学習率(DDCLR)は、実・混合データセット双方に対して低いチューニングコストで汎用的に高精度な学習結果を提供する点において有用性があると考えられる。実データセットに対しては表 3, 5 より DDCLR, CLR, DCLR(1)は表 1 に示すその他の学習率設定と比較し、基本的に一定水準以上の精度を 5 つのデータセットに対して示している。一方、DCLR(2)の精度が悪くなっており、DCLR(3)もデータセットによって精度の良し悪しは分かれ、DCLR の中で初期設定による結果のばらつき

が見られる結果となった。このことより、周期性をもつ学習率設定は実データセット学習において有効性が比較的高いが、減衰周期性は初期値依存性が高いと考えられる。一方で全ての学習率設定における学習精度結果は一定程度の成果が得られたことより、結論として提案手法は実データに対して少なくとも他の学習率設定と同等の学習結果が得られるといえる。混合データセットに対しては、表 4, 6 より DDCLR が他の学習率設定に比べて群を抜いて精度が高いことがわかる。オプティマイザの Adam で学習率チューニングを行っている Fixed や CLR は精度を大きく落とし、減衰性、下降性を持つものの精度が比較的高いことがわかる。しかしながら DCLR と ExpLR は初期設定依存性による結果の良し悪しが現れており、汎用性に欠けることから高いチューニングコストが求められる。このような現象は、実・混合データセット学習の一般化においてブラックボックスな状態のモデル利用者にとって好ましくないものである。DDCLR は実・混合データセットのそれぞれ 5 つのデータセットに汎用的に低いチューニングコストで高い学習精度を達成できた点において有用性が高いと考える。

## 5. 結論

### 5.1 まとめ

本研究で述べた実・混合データセット双方の学習の一般化に対する課題点は 2 つである。1 つ目は学習率設定における一部のケースで、実データセット学習時に有用性を示した学習率設定であっても、混合データセット学習時に他の設定に比べ大幅に精度を低下させる危険性があるという点、2 つ目は実・混合データセット学習の一般化により、各データセットの学習に汎用的に機能するような学習率設定がより難しくなり、チューニングコストを増大させる、または予期せぬ品質低下を招くという点である。

以上の課題に対し、本研究で提案する下降型減衰周期型学習率の導入によって、低いチューニングコストで、実・混合データセット双方に対して高精度かつ汎用的に学習をサポートできる可能性があることを示した。

### 5.2 今後の取り組み

Stylegan2 のみならず、その他の生成モデルに対しても汎用的に実・混合データセット双方の学習を実現するための研究を活性化させる必要があると考える。そのために、学習率 Scheduler に限らず、その他の着眼点から総合的に解決を目指していきたい。

### 参考文献

- [1] Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luz, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkooh, Richard G. Baraniuk, "Self-Consuming Generative Models Go MAD", ICLR, 2024.
- [2] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila, "Analyzing and Improving the Image Quality of StyleGAN", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8110-8119.
- [3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, "High-Resolution Image Synthesis with Latent Diffusion Model", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684-10695.
- [4] Leslie N. Smith, "Cyclical Learning Rates for Training Neural Networks", IEEE Winter Conference on Applications of Computer Vision (WACV) (2017).

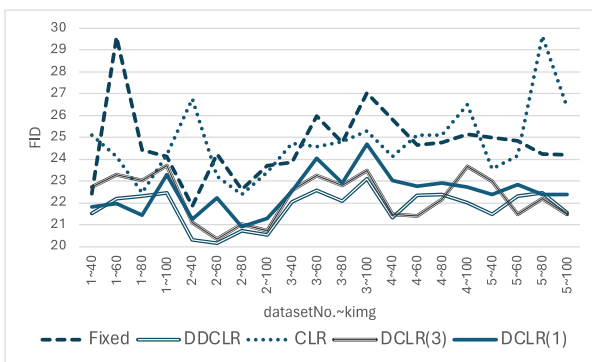


図 5 Mix-dataset 転移学習時の FID 推移