機械知覚&ロボティクスグループ
中部大学　工学部　ロボット理工学科

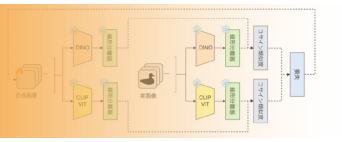
MPRG
MACHINE PERCEPTION AND ROBOTICS GROUP

2025年度　藤吉研究室　卒業論文発表　アブストラクト

Dataset Distillation, Linear Gradient Matching (LGM)

複数モデルの勾配を統合したデータセット蒸留による下流タスクの高精度化

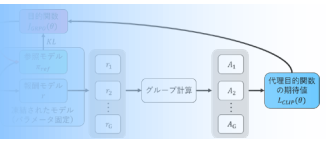
市川 翔



Imitation Learning, Reinforcement Learning

拡散モデルにおける模倣学習の動作保持と環境適応を両立した強化学習

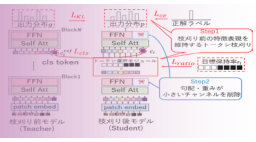
児玉 将汰



Pruning, Distillation

特徴表現の維持に着目した構造化枝刈りとトークン枝刈りの併用

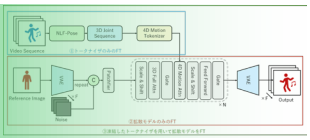
篠田 昂希



Diffusion Model, Fine-tuning

スポーツ動作に対応した MTVCrafter のファインチューニング戦略

野田 悠太



Packing Rate, BEV

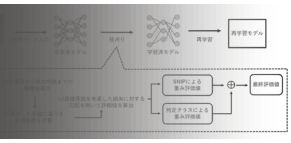
サッカー戦術分析のためのパッキングレート自動算出システム

古田 愛貴

Pruning, Object Dection, BEV

精度劣化を抑制した BEVFormer の重みつき非構造枝刈り

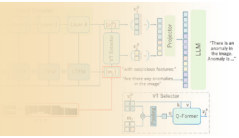
松本 大輝



Multimodal Large Language Model, Look-Twice Feature Matching, Visual Token Selector

マスク誘導型 Visual Encoder による異常検知の言語説明と可視化

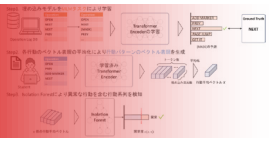
森田 悠斗



Isolation Forest, Behavioral Anomaly Detection, Learning Behavior Patterns

学習行動の文脈を考慮した教師なし異常行動検知

安田 恭大



URL: http://mprg.jp

複数モデルの勾配を統合したデータセット蒸留による下流タスクの高精度化

ER22005 市川翔　　指導教授：藤吉弘亘

1. はじめに

Contrastive Language-Image Pre-Training (CLIP) や Distillation with no labels (DINOv1) に代表される事前学習済みモデルは，下流タスクでの評価において高い性能を示しており，多様な画像認識タスクの基盤として広く利用されている．事前学習済みモデルを下流タスクへ転移学習をする際，計算コストを抑制するために限られた学習データによる効率的な学習が重要である．その 1 つのアプローチとして学習データを少数の合成データに凝縮するデータセット蒸留が注目されている．従来のデータセット蒸留手法は単一モデルに基づく手法であり，学習目的の違いに起因する特徴の多様性を同時に反映できない可能性がある．そこで本研究では，Linear Gradient Matching [1] を拡張し，特性の異なる複数の事前学習済みモデルを用いたデータセット蒸留手法を提案する．本手法により，未知のモデルに対しても分類に有効な特徴を保持する合成画像の作成を目指す．

2. Linear Gradient Matching

データセット蒸留の代表的な手法として，Linear Gradient Matching (LGM) [1] が提案されている．LGM は，事前学習済みモデルを固定した状態で，実画像と合成画像に対する線形分離器の勾配が一致するように合成画像を最適化する．これにより，分類に有効な特徴を合成画像に凝縮できる．しかし，LGM は単一モデルの勾配に基づいて最適化を行うため，得られる合成画像は蒸留に使用したモデルの特徴表現に依存する．その結果，異なるアーキテクチャを持つモデルに対しては，転移性能が十分に発揮されないという課題がある．

3. 提案手法

本研究では図 1 に示すように，LGM を拡張し，複数の事前学習済みモデルを同時に用いるデータセット蒸留手法を提案する．学習目的の異なるモデルの勾配を同時に用いることで，特定のモデルに依存しない合成画像の作成を目的とする．本研究では，自己教師あり学習に基づく DINO 系モデルと，画像と言語対照学習に基づく CLIP 系モデルを組み合わせる．提案手法における勾配損失を式 (1) に示す．本手法では，式 (1) を最小化するように，合成画像の作成のパラメータを最適化する．具体的には，事前学習済みモデルおよび線形分離器のパラメータは学習せずに，誤差逆伝播法により合成画像に対する線形分離器の損失勾配を計算し，実画像から得られる勾配との類似度が高くなるように，合成画像を更新する．ここで，g^{DINO} および g^{CLIP} は，それぞれ DINO 系モデルおよび CLIP 系モデルで得られる線形分離器の勾配を表す．

$$\mathcal{L}_{\mathrm{grad}} = (1 - \cos(g^{\mathrm{DINO}})) + (1 - \cos(g^{\mathrm{CLIP}})) \qquad (1)$$

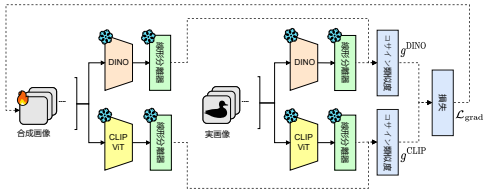


図 1: 提案手法の概要

4. 評価実験

本研究では，作成した合成画像が蒸留に使用していないモデルに対しても分類に有効な特徴表現を保持しているかを検証するため，事前学習済みモデルのバックボーンを凍結し，最終層の線形分離器のみを学習する Linear Probing を用いて分類精度評価を行う．評価モデルには，蒸留に使用していない CNN アーキテクチャである ResNet50 を用いた．評価には 20 クラスの画像を用い，各クラス 5 枚の画像から線形分離器を学習した．実験は 5 回実施し，その平均値と標準偏差を評価値とした．また，比較として全学習データを用いた場合の精度を上限値として併記する．画像の作成には，DINOv1，DINOv2，CLIP，Sigmoid Loss for Language Image Pre-training（SigLIP）の単体モデルおよびそれらの組み合わせを用いた．

4.1. 定量的評価

分類精度の評価結果を表 1 に示す．単一モデルの場合と比較して，複数モデルを統合した場合では，多くの組み合わせにおいて分類精度の向上が確認できる．具体的に，DINOv2 と SigLIP の組み合わせを DINOv2 単体と比較した場合を除き，精度の向上が確認できる．さらに，複数モデルを統合した場合，DINOv2 と SigLIP を組み合わせた場合を除いて，実画像を用いた場合を上回る精度となった．また，単一モデルでは分類精度が低かった SigLIP においても，他のモデルと統合することで分類精度の向上がみられた．この結果から，複数モデルの勾配を統合することで，分類に有効な特徴をより効果的に合成画像へ反映することができることがわかった．

表 1: Top1 Accuracy [%]

画像の種類	画像作成に使用したモデル				評価モデル
	DINO 系		CLIP 系		ResNet50
	v1	v2	CLIP	SigLIP	
上限値	–	–	–	–	98.46 ± 0.05
実画像	–	–	–	–	95.86 ± 0.05
合成画像 (単一モデル)	✓	–	–	–	96.24 ± 0.19
	–	✓	–	–	95.68 ± 0.15
	–	–	✓	–	92.28 ± 0.12
	–	–	–	✓	72.38 ± 0.34
合成画像 (複数モデル)	✓	–	✓	–	96.68 ± 0.07
	✓	–	–	✓	97.06 ± 0.15
	–	✓	✓	–	96.68 ± 0.24
	–	✓	–	✓	95.68 ± 0.07

4.2. 定性的評価

図 2 にフラミンゴの合成画像例を示す．単一モデルによる合成画像では，モデルごとに着目する視覚的要素が異なる．図 2 (a) の DINOv2 では，フラミンゴのシルエットや形状が強調されている．一方，図 2 (b) の CLIP では，背景の水面などのテクスチャ表現が強調されている．これに対し，図 2 (c) の DINOv2 と CLIP の組み合わせでは，形状とテクスチャ情報が含まれているようにも見受けられるが，単一モデルと比較して視覚的な差異は明確ではない．この結果は，分類に有効な特徴が，必ずしも視覚的に識別可能な形で現れないことを示唆している．

(a) DINOv2　(b) CLIP　(c) v2+CLIP

図 2: 作成した合成画像

5. おわりに

本研究では，LGM を拡張し，複数の事前学習済みモデルの勾配を同時に最適化する LGM を提案した．定量的評価により，提案手法で作成した合成画像が，単一モデル蒸留と比較して高い分類精度を示すことを確認した．この結果は，提案手法が複数の事前学習済みモデルの特徴を統合し，分類に有効な特徴表現を抽出可能であることを示唆している．今後の展望として，モデル統合時の各バックボーンの重み最適化や中間層の特徴活用による合成画像の質向上，および対象クラス数やデータセットの拡大が挙げられる．さらに，Linear Probing 以外の手法への適用における課題と可能性についても，検証を行う予定である．

参考文献

[1] G. Cazenavette, *et al.*, “Dataset Distillation for Pre-Trained Self-Supervised Vision Models”, NeurIPS, 2025.

拡散モデルにおける模倣学習の動作保持と環境適応を両立した強化学習

ER22027 児玉将汰　　指導教授：藤吉弘亘

1. はじめに

模倣学習は人の動作を教師として模倣するように学習するため，未知の状況下では適切な動作ができない課題がある．この課題に対して強化学習による方策の更新が有効だが，強化学習することで模倣学習により獲得した動作を忘却するという問題がある．本研究では，模倣動作を保持しつつ強化学習する手法を提案する．提案手法により，動作の忘却を低減しつつ環境適応を図ることを目的とする．

2. 従来手法

模倣学習したモデルの重みを初期値として，強化学習により，方策を更新する DPPO [1] が提案されている．DPPO は，Diffusion Policy [2] の多様な動作候補を生成できる能力と強化学習を組み合わせることで，把持に失敗してもその位置を再認識し，正しい場所に戻す動作の自律的な学習が可能である．しかし，方策を逐次更新するため，方策が学習の進行で徐々に変化する．結果として，模倣学習で獲得した動作を忘却することがある．

3. 提案手法

本研究では，DPPO [1] に最適化手法 GRPO(Group Relative Policy Optimization) [3] を導入することで模倣学習で獲得した方策を参照モデルとして忘却をしないように方策の更新を行う手法を提案する．GRPO は複数の動作候補を生成し，グループ内での相対的な報酬で学習する手法であり，多様な動作候補の中から適切な動作を探索可能である．学習の際には，ResNet18 で抽出したカメラ画像の特徴量と関節の状態を入力し，関節の目標角度を出力する．GRPO による方策更新の手順を図 1 に示す．

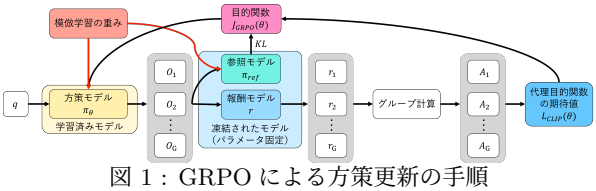


図 1：GRPO による方策更新の手順

黒色の矢印は学習中のデータの流れや演算で，赤色の矢印はパラメータの継承（初期値）を表す．まず，方策モデルが入力 q に対して複数の動作候補 O のグループを生成する．次に，報酬モデルで各応答を評価して報酬 r を算出した後，グループ内での相対的な優劣を示すアドバンテージ A を決定する．学習には，式 (1) に示す目的関数 $J_{GRPO}(\theta)$ を用いる．

$$J_{GRPO}(\theta) = \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\mathcal{L}_{CLIP}(\theta)\right] - \beta\mathcal{D}_{KL}(\pi_\theta||\pi_{ref}) \tag{1}$$

$\mathcal{L}_{CLIP}(\theta)$ はアドバンテージに基づく代理目的関数の期待値を表し，π_{ref} は模倣学習の重みを継承しパラメータを固定した参照モデルである．学習中の方策モデル π_θ と参照モデル π_{ref} の KL ダイバージェンスを算出しており，これにより π_θ が模倣学習時の動作から過度に乖離することを抑制する．これにより，強化学習による未知環境への適応を促進しつつ，模倣学習で獲得した基本動作の維持を図る．

4. 評価実験

本実験では模倣学習と 2 種類の強化学習（PPO/GRPO）によるロボットの動作制御を行い，環境の差異が与える影響や，軌道生成の評価を行う．

4.1. 実験概要

Genesis シミュレータ内でヒューマノイドロボットの Unitree G1 を用いた実験を行う．両腕 28 関節を制御対象とし，物体把持タスクの模倣学習と強化学習を行う．使用するデータセットは実機の Unitree G1 で収集されたブロック積み上げタスクのデータであり，頭部 2 視点と両手カメラの計 4 視点の RGB 画像から構成される．強化学習の報酬は，把持対象への接近・接触および，把持の成功だけを条件にした簡易的なものとし，模倣学習による事前学習動作を活用しやすくする．

4.2. 実験条件

学習では共通して，方策モデルに Diffusion Policy，最適化手法に AdamW，学習回数を 1000 として学習する．模倣学習では，学習手法を Behavior Cloning，バッチサイズを 8，学習率を 1e-4 として学習する．強化学習では，学習手法を PPO，バッチサイズを 256，学習率を 1e-4 として学習する．提案手法は，学習手法を GRPO，バッチサイズを 32，学習率を 1e-6 として学習をする．模倣学習ではオープンソースの G1_Dex3_BlockStacking_Dataset をデータセットとして使用する．強化学習では容量削減や学習速度向上のため，カメラの画質を最低限にして実行する．

4.3. 実験結果

従来手法を青線，提案手法を赤線として学習中の報酬の推移を図 2 に示す．

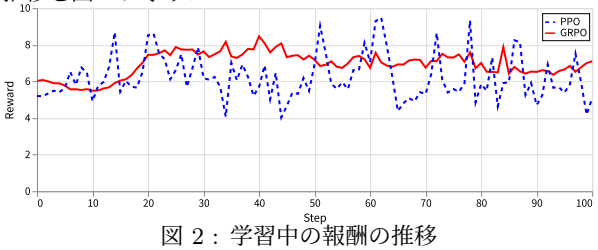


図 2：学習中の報酬の推移

従来手法は報酬の変動が激しく，提案手法は安定している．従来手法および提案手法におけるロボット頭部のカメラ視点の描画結果を図 3，図 4 に示す．模倣学習のみでは，把持対象を認識して，腕を伸ばす動作や，把持対象付近で指を曲げて，腕を上げる動作はできたが，把持には至らなかった．モデルを対象とする従来手法は，把持動作を確認（青丸）できたが，学習を進めるほど事前学習で獲得した動作から逸脱する傾向が見られた．動作が不自然で，別のブロックへの接触（黄丸）や，右腕の画面外への移動（赤丸）等の動作が見られた．一方，提案手法は，従来手法よりも事前学習に沿った動作が維持され，把持対象に充分接近してから指を曲げる動作（紫丸）が確認できた．従来手法よりも安定した動作でありながら，模倣学習よりも確実に把持対象を認識して近づく傾向（緑丸）が見られた．

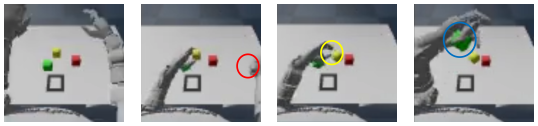


図 3：従来手法の把持動作

図 4：提案手法の把持動作

5. おわりに

本研究では，模倣学習後の強化学習によるロボットの動作精度の向上の手法の提案した．提案手法によって，模倣学習で獲得した動作を保持しつつ，強化学習が可能だと確認できた．今後は，提案手法による物体把持後の積み上げタスクを行う予定である．

参考文献

[1] AZ. Ren, *et al.*, “Diffusion Policy Policy Optimization”, ICLR, 2025.

[2] C. chi, *et al.*, “Diffusion Policy: Visuomotor Policy Learning via Action Diffusion”, RSS, 2023.

[3] Z. Shao, *et al.*, “DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models”, arXiv, 2024.

特徴表現の維持に着目した構造化枝刈りとトークン枝刈りの併用

ER22034 篠田昂希　　指導教授：藤吉弘亘

1. はじめに

Vision Transformer（ViT）による物体認識モデルは高精度化に伴いパラメータ数が大規模化している．そのため，エッジデバイス適用やリアルタイム推論には，計算量とメモリ使用量が課題である．この課題に対して，入力トークン数を削減し，モデル全体の計算量を抑えるトークン枝刈りが注目されている．トークン枝刈りは入力を削減することで計算量を減らせるため，推論速度向上に有効である．一方で，トークンを過度に削減すると特徴表現が不足し，精度が著しく低下する．また，モデルのパラメータ数は変化しないため，モデル軽量化が不可能である．そこで本研究では，トークン枝刈りと構造化枝刈りを併用し，特徴表現と精度の維持を図りながら推論速度向上とモデル軽量化の両立を目指すハイブリッド手法を提案する．

2. トークン枝刈り

トークン枝刈りの代表的手法として DynamicViT[1] がある．DynamicViT は ViT のエンコーダ間にトークン選択モジュールを挿入し，各トークンの重要度に基づいて残すトークンを決定する．学習時は Gumbel-Softmax により選択をスコア化して学習可能にし，推論時は重要度の低いトークンを破棄して計算量を削減する．DynamicViT では，各層の保持率が目標値に近づくよう制約を与える損失 L_{ratio} を導入し，式 (1) で定義する．ここで $\mathcal{S}$ はトークン選択モジュールを挿入した層の集合，l はその層の添字である．また，r_l は目標保持率，$\hat{r}_l$ は実際の保持率である．

$$L_{\text{ratio}} = \frac{1}{|\mathcal{S}|} \sum_{l \in \mathcal{S}} |\hat{r}_l - r_l| \tag{1}$$

また，教師モデルの出力分布 $\mathbf{q}$ と生徒モデルの出力分布 $\mathbf{p}$ の差を抑える蒸留損失 L_{KL} を導入し，式 (2) で定義する．ここで k はクラスの添字であり，q_k および p_k はクラス k に対応する確率である．

$$L_{\text{KL}} = \sum_{k} q_k \log \frac{q_k}{p_k} \tag{2}$$

3. 提案手法

提案手法は図 1 に示すように，2 段階で枝刈りを行う．Step1 では cls token の特徴表現整合を導入したトークン枝刈りを行い，Step2 では勾配に基づいて MLP チャンネルを枝刈りする．

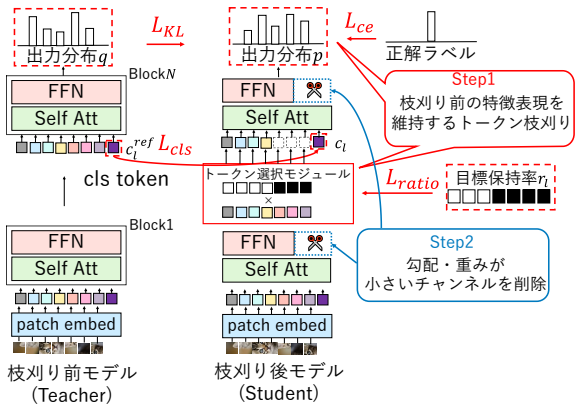


図 1：2 段階の枝刈り

3.1. cls token のコサイン類似度に基づく損失導入

Step1 では，DynamicViT に枝刈り前後の cls token のコサイン類似度に基づく損失 L_{cls} を追加する．

$$L_{\text{cls}} = \frac{1}{|\mathcal{S}|} \sum_{l \in \mathcal{S}} \left(1 - \cos(\mathbf{c}_l, \mathbf{c}_l^{\text{ref}})\right) \tag{3}$$

式 (3) に cls token の特徴表現に基づく損失項を表す．ここで $\mathbf{c}_l$ は Student の cls token，$\mathbf{c}_l^{\text{ref}}$ は Teacher の cls token である．総損失 L を式 (4) に示す．

$$L = L_{\text{ce}} + \lambda_{\text{ratio}} L_{\text{ratio}} + \lambda_{\text{KL}} L_{\text{KL}} + \lambda_{\text{cls}} L_{\text{cls}} \tag{4}$$

ここで L_{ce} はクロスエントロピー損失であり，$\lambda_{\text{ratio}}, \lambda_{\text{KL}}, \lambda_{\text{cls}}$ はハイパーパラメータである．

3.2. MLP 構造化枝刈り（勾配ベースの重要度推定）

Step2 では，式 (5) のようにトークン枝刈り学習の損失に対する重みの勾配を用いて MLP チャンネルの重要度を推定する．重要度の小さい順に枝刈りする．重要度は一次テイラー近似に基づき，重みと勾配をチャンネルごとに集約して算出する．ブロック N の MLP において，チャンネル c の重要度 $s_{n,c}$ を式 (5) により定義する．

$$s_{n,c} = \sum_{j} \left| \frac{\partial L}{\partial W_{c,j}^{(n)}} \cdot W_{c,j}^{(n)} \right| \tag{5}$$

ここで $W_{c,j}^{(n)}$ は MLP の重みであり，$s_{n,c}$ が小さいチャンネルほど出力への寄与が小さいとみなして枝刈りする．

4. 評価実験

計算量を揃えた条件で，枝刈り手法ごとの正解率，Throughput，および cls token の変化を比較する．

4.1. 実験概要

ImageNet-1k で事前学習済みの DeiT-Base/16 を用い，下流タスクとして CIFAR100 および StanfordDogs で枝刈り・評価する．比較手法は，ベースライン（枝刈り前モデル），Magnitude（MLP の Magnitude 構造化枝刈り），DynamicViT，単純併用（DynamicViT + Magnitude）である．推論時の計算量が約 8.0 GFLOPs となるように枝刈り率を調整する．トークン枝刈りはエンコーダ層 $\{3, 6, 9\}$ に適用する．MLP 構造化枝刈りは MLP チャンネルを削減する．評価指標は正解率，Throughput，および枝刈り前モデルを参照した cls token の特徴表現の変化とする．

4.2. 実験結果

表 1 に，各手法を約 8.0 GFLOPs に揃えた条件での性能を示す．ベースラインと比較すると，提案手法は計算量を抑えつつ推論速度を向上させ，精度低下も小さい．単純併用は高速化できる一方で精度低下が大きく，Magnitude 基準ではトークン枝刈り後の中間層の出力の分布変化を反映できないため，特徴表現が劣化すると考えられる．

表 1：CIFAR100 および StanfordDogs における性能比較

Method	CIFAR100		StanfordDogs	
	Acc↑	Thr.↑	Acc↑	Thr.↑
ベースライン	89.70	318.00	95.08	291.21
Magnitude	70.21	683.79	53.81	558.31
DynamicViT	80.58	**843.02**	80.67	**783.76**
単純併用	80.06	807.54	80.41	758.58
提案手法	**87.19**	808.86	**93.67**	761.27

表 2 に cls token の特徴表現の変化を示す．提案手法は単純併用と比べて特徴表現の変化が小さい傾向が確認でき，枝刈り前に近い特徴表現を維持できたといえる．

表 2：枝刈り前後の cls token の特徴表現の変化

Method	CIFAR100		StanfordDogs	
	cos sim↑	L1Norm↓	cos sim↑	L1Norm↓
Magnitude	0.699	1693.5	0.583	1773.53
DynamicViT	0.760	1554.0	0.753	1545.76
単純併用	0.403	2859.8	0.742	1631.72
提案手法	**0.980**	**697.33**	**0.869**	**777.16**

5. おわりに

本研究では，トークン枝刈りと MLP 構造化枝刈りを 2 段階で適用し，特徴表現を維持しながら高速化する手法を提案した．CIFAR100 および StanfordDogs において，提案手法は単純併用より特徴表現の変化が小さく，精度低下を抑えられることを確認した．

参考文献

[1] Y. Rao, *et al.*, “DynamicViT: Efficient Vision Transformers with Dynamic Token Sparsification”, NeurIPS, 2021.

スポーツ動作に対応した MTVCrafter のファインチューニング戦略

ER22051 野田悠太　　指導教授：藤吉弘亘

1. はじめに

スポーツにおいて, 映像フィードバックは選手のパフォーマンス向上に有効である．フィードバックの際は, 選手の動作を任意の条件で再現可能な映像生成が求められる．しかしながら, スポーツ特有の急激な動作に対応した映像生成の実現には, 対象とする動作に特化した学習データが必要となる．拡散モデルを活用した生成モデルである MTVCrafter は，web 上から収集した日常生活の映像を用いて学習されており，スポーツ特有の動作の生成能力は不十分である．そこで本研究では，スポーツ領域での映像生成の品質向上を目指して，MTVCrafter のファインチューニング戦略を検討する．

2. MTVCrafter

MTVCrafter[1] は, 参照画像と 3 次元関節座標の系列データから参照画像に映る人物の動作映像を生成する．MTVCrafter のモデル構造を図 1 に示す．MTVCrafter は, VAE を用いて画像を潜在空間へ写像し，潜在特徴を獲得する．そして，潜在特徴と動作情報を Transformer に入力して動画を生成する．その際，Cross-Attention を通じて動作情報を条件として取り込む．この時，学習していない動作情報は，適切な条件とならないため，破綻した映像が生成されることがある．

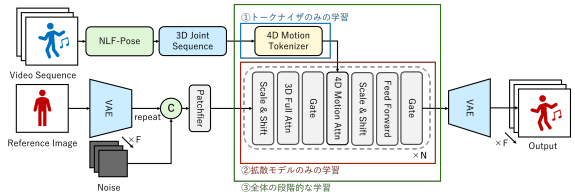


図 1: MTVCrafter のモデル構造

3. 提案手法

本研究の目的は，MTVCrafter のスポーツ領域での映像生成の品質向上である．そのため, モデルをファインチューニングするための学習戦略を検討する．具体的には, スポーツ特有の急激な動作としてフィギュアスケートのジャンプに着目し，後述の独自データセット FSJD を用いたファインチューニングを行う．

3.1. 学習戦略

MTVCrafter の学習には，①トークナイザのみの学習，②拡散モデルのみの学習，③全体の学習，の 3 つの過程がある．各学習過程が生成映像に与える影響を比較検証する．

3.2. 動画-3D ポーズ系列ペアのデータセット作成

本研究では，Figure Skating Jump Dataset (FSJD) を作成した．まず，収集した 272 本のフィギュアスケートの競技映像からジャンプシーンを切り出した．次に，各シーンに対して 3 次元人体形状・姿勢推定モデル NLF-Pose [2] を用いて，シーンの各フレームに 3 次元関節座標を付与した．得られた FSJD のデータ数は 1,072 組で，各データは解像度 512 × 512 画素，フレーム数 49 に統一されている．

4. 評価実験

本実験では，以下の学習戦略の異なる 3 つのモデルの生成性能を比較する．

モデル 1 ：トークナイザのみの学習

モデル 2 ：拡散モデルのみの学習

モデル 3 ：トークナイザを学習・凍結後, 拡散モデルを学習

本実験では, FSJD のうち学習用に 972 組, 評価用に 100 組を使用した．

4.1. 各学習戦略におけるトークナイザの比較

トークナイザの実験結果を表 1 に示す．トークナイザを学習した場合，しない場合と比較して MPJPE の低下からトークン化の再現性が確認できる．また，FID の大幅な低下から FSJD 特有の動作分布を適切に学習したといえる．

表 1: 各学習戦略の比較

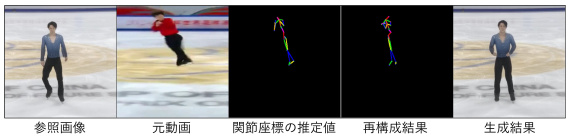
評価指標	モデル 1	モデル 2	モデル 3
MPJPE	287.28	302.47	287.28
FID	186.17	2798.47	186.17

4.2. 各学習戦略における定性的比較

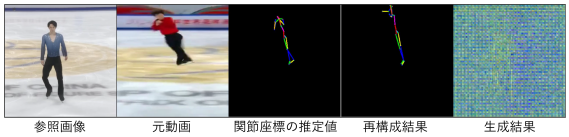
各学習戦略における推論例の比較を図 2 に示す．図 2(a) より，トークナイザのみを学習した場合，入力された動作条件を無視し，学習済みの一般的な動作が生成された．これは，拡散モデルの Cross-Attention 層が未知ドメインのトークンに対応しておらず，条件情報が意味を持たないノイズとして処理されたためと考えられる．

図 2(b) より，拡散モデルのみを学習した場合，不定形の青いノイズが生成された．これは，未学習の急激な動作に対して，トークナイザが適切に情報を保ってトークン化できず，正解の映像と矛盾した条件情報が入力されたことで，拡散モデルの最適化が阻害されたためと考えられる．

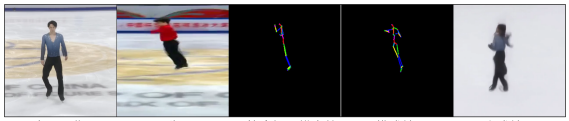
図 2(c) より，全体を段階的に学習した場合，既存モデルでは生成に失敗しやすかった空中での急激な回転時に，外観や動作が失敗なく生成されている．そのため，全体の学習がドメイン特有の動作再現において有効といえる．



(a) トークナイザのみを学習した際の推論例



(b) 拡散モデルのみを学習した際の推論例



(c) 全体を学習した際の推論例

図 2: 各学習戦略における推論例の比較

4.3. 全体学習の定量的評価

全体を段階的に学習した結果を表 2 に示す．特に FVD 及び FID-VID の改善から，ドメイン特化の学習により，ジャンプ特有の動作に対応した生成が可能となった．

表 2: 各評価指標における結果比較

比較対象	FVD	FID-VID	PSNR	SSIM	LPIPS	FID
学習前	479.15	39.11	13.10	0.457	0.525	35.67
学習後	303.39	35.30	13.58	0.577	0.530	35.73

5. おわりに

本研究では，MTVCrafter の学習データにない別ドメインでの映像生成の品質向上を試みた．具体的には，FSJD を構築し，トークナイザ及び拡散モデルをファインチューニングした．結果，全体を段階的に学習した場合に，定性・定量評価において，既存モデルでは難しかったジャンプ特有の動作の再現性の向上といった映像生成の品質向上を確認した．今後は，さらにテキスト情報を条件として加えることで，生成における能動的なジャンプ種別の指定を図る．

参考文献

[1] Yanbo Ding, *et al.*, “MTVCrafter: 4D Motion Tokenization for Open-World Human Image Animation”, arXiv:2505.10238, 2025.

[2] István Sárándi and Gerard Pons-Moll, “Neural Localizer Fields for Continuous 3D Human Pose and Shape Estimation”, arXiv:2407.07532, 2024.

サッカー戦術分析のためのパッキングレート自動算出システム

ER22063 古田愛貴　　指導教授：藤吉弘亘

1. はじめに

サッカーの試合では，シュート数やパス数などの定量指標に基づく分析が行われている．パスの質を定量化する新たな指標として，パッキングレートが注目されている．パッキングレートは，1 本のパスやドリブルによって攻撃方向に位置する相手選手を何人通過したかを定量的に評価できる．これまで，パッキングレートの算出は試合映像を目視で確認する方法が主流で，算出に多大な時間と人的コストを要するという課題があった．そこで本研究では，試合映像からパッキングレートを自動で算出する手法を提案する．

2. 従来手法

従来のパッキングレート算出では，攻撃方向に基づく一次元的な選手の位置関係によって評価が行われる．そのため，図 1 に示すシーンでは，A や E のように，パスに直接関与しない選手まで評価対象に含まれ，パッキングレートは 5 と算出される．

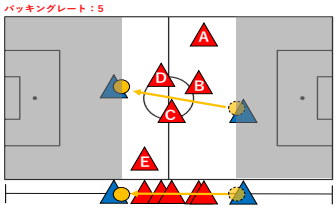


図 1：パッキングレートイメージ図

また，算出の際は試合映像を目視で確認し，相手選手数を手作業でカウントするため，時間と労力を要する．

3. 提案手法

本研究では，試合映像から取得したコート，選手，ボールの情報を用い，パッキングレートを自動的に算出する手法を提案する．

3.1. 処理概要

提案手法では，試合映像から YOLOv8[1] を用いて，コート，選手，ボールの位置を検出する．選手は検出後，色情報をもとにチームを判別する．コートの基準点はペナルティエリア，センターラインなどのコーナー 32 点を検出し，検出したコートライン情報を用いて Bird's Eye View（BEV）視点へ変換する．これにより，距離歪みを軽減した BEV 座標上で，ボール周囲の一定範囲内に存在する守備側選手のみを対象としたパッキングレートの算出可能となる．処理フローを図 2 に示す．

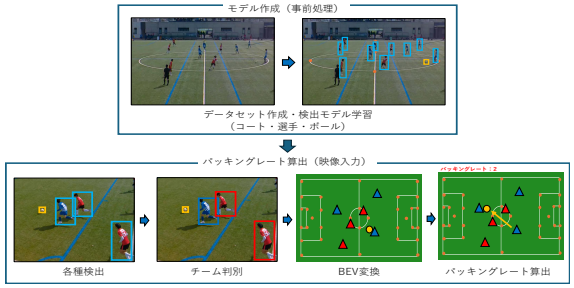


図 2：提案手法の処理フロー

本研究では従来研究に倣い，パッキングレートを 1 プレーで通過した相手選手人数をポイントとして表した指標と定義し，以下の式で表す．

$$\mathrm{P} = \sum_{i \in \mathcal{D}} \delta(i) \tag{1}$$

ここで，$\mathcal{D}$ は BEV 座標上においてボール周囲の一定範囲内に存在する守備側選手の集合を表す．$\delta(i)$ は，パスまたはドリブル区間において BEV 上でボールが守備側選手 i を通過した場合に 1，それ以外を 0 とする指示関数である．

3.2. データセット作成

データセット作成および学習には，Roboflow[2] を用いた．3 試合分，計 4,137 枚の画像に対して選手，ボール，コートのコーナー点のアノテーションを行った．

3.3. チーム判別

作成した選手検出モデルの結果からチーム判別を行う．各選手のバウンディングボックス領域を入力とし，事前学習済みの画像特徴抽出モデルを用いて選手画像の特徴量を抽出する．抽出された高次元特徴量に対し，UMAP を用いて 3 次元へ次元削減を行い，k-means 法（$k = 2$）によるクラスタリングを適用することで，2 チームに分類する．

3.4. BEV 視点への変換

カメラ映像上の座標は，遠近効果により距離関係が歪むため，正確な位置関係を扱うことが困難である．そこで，学習したコートライン検出モデルにより取得した映像座標と BEV 座標のキーポイント対応を用いてホモグラフィ変換を行い，映像を BEV 視点へ変換する．さらに，追跡安定化のためカルマンフィルタを適用する．

4. 評価実験

4.1. 実験概要

本研究では，ボール位置に追従してカメラが左右にパンするサッカー試合映像を対象とする．提案手法の有効性を検証するため，実際のサッカー試合映像からパス，ドリブル 10 シーンを抽出し評価を行った．各シーンに対して，BEV 視点を用いない画像座標上で算出する手法と，BEV 視点を用いた提案手法の 2 つの手法を比較し，目視結果と一致した場合を算出成功と定義した．

4.2. 実験結果

提案手法による BEV への変換結果を図 3 に示す．図 3 より，選手およびボールの位置関係が正しく BEV 視点に投影されていることが分かる．

図 3：BEV 視点への変換結果

次にパッキングレートの算出結果を表 1 に示す．表 1 より，画像座標ベース手法では真値に対して最大で 9 人の誤差が生じたのに対し，BEV 変換手法では誤差が小さく，一致率が 60%に向上した．これは，BEV 視点を用いることで距離歪みの影響が軽減され，選手とボールの位置関係をコート平面の距離として評価可能になったためである．

表 1：パッキングレート数精度比較

シーン	真値	BEV 変換手法		画像座標ベース手法	
		測定値	一致	測定値	一致
1	3	3	✓	5	-
2	2	2	✓	2	✓
3	3	3	✓	7	-
4	2	2	✓	6	-
5	3	1	-	8	-
6	1	1	✓	4	-
7	2	2	✓	5	-
8	3	6	-	8	-
9	1	2	-	6	-
10	1	4	-	10	-
一致率		6/10		1/10	

5. まとめ

本研究では，試合映像からパッキングレートを自動で算出する手法を提案した．評価実験の結果，BEV 視点を用いることで，算出成功率が向上した．一方で，カメラ移動によりコート基準点の検出が不十分なフレームでは，BEV 変換が不安定となる課題が残った．今後は，BEV 変換の安定化を図り，試合全体を通したパッキングレートの自動算出を可能とする．

参考文献

[1] G. Jocher, A. Chaurasia, and J. Qiu, " YOLO by Ultralytics ", arXiv preprint arXiv:2301.07209, 2023.

[2] J. Nelson, J. Solawetz, and J. Houghton, "Roboflow: Simplified Computer Vision Model Training and Deployment", Roboflow Inc., 2023.

精度劣化を抑制した BEVFormer の重み付き非構造枝刈り

ER22070 松本大輝　　指導教授：藤吉弘亘

1. はじめに

自動運転車が安全に走行するには，自車周辺の物体の 3 次元位置情報が必要である．その代表的な手法として 3D 物体検出が用いられるが，高性能化に伴いモデルの大規模化が課題となっている．この課題を解決する手法として，モデルをコンパクト化する枝刈りがある．枝刈りは構造化枝刈りと非構造枝刈りの 2 つに大別される．非構造枝刈りは重み行列を要素単位で削除することで，ネットワーク構造を維持したまま大幅にパラメータを削減できるという利点がある．この利点を踏まえ，本研究では非構造枝刈りの物体検出モデルへの有効性を検証するとともに，枝刈りによる精度低下を抑制するために特定クラスおよび距離の重み付けを導入した非構造枝刈りを提案する．

2. 先行研究

非構造枝刈りの評価値として SNIP が用いられる．SNIP は損失関数に影響を与える重みを削除する手法であり，その評価値は，損失関数を $\mathcal{L}$，i 番目の重みを w_i として $S_{\mathrm{SNIP}}(w_i) = \left|\frac{\partial \mathcal{L}}{\partial w_i} w_i\right|$ のように示す．SNIP による枝刈りでは特定クラスの検出精度が低下することがある．先行研究 [1] では，特定クラスの精度低下を抑制するため，特定クラスを含む入力 x^{specific} に対する損失から $S_{\mathrm{class}}(w_i) = \left|\frac{\partial \mathcal{L}(x^{\mathrm{specific}})}{\partial w_i} w_i\right|$ のように示す．そして，S_{SNIP} と S_{class} を統合し枝刈りを行う．最終評価値 $S(w_i)$ を式 (1) に示す．

$$S(w_i) = S_{\mathrm{class}}(w_i) + S_{\mathrm{SNIP}}(w_i) \tag{1}$$

3. 提案手法

自動運転の観点では近距離領域での検出失敗が交通事故に直結するため，特定クラスに加え，近距離での検出精度低下を抑制する必要がある．そこで本研究では代表的な 3D 物体検出モデルである BEVFormer[2] を対象に，精度低下を抑制する特定クラスおよび距離の重み付けを導入した非構造枝刈りを提案する．提案手法の概要を図 1 に示す．

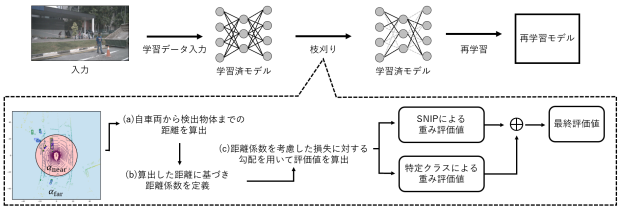


図 1：提案手法の概要

提案手法では，距離係数 $\alpha(d)$ を付与した損失関数 $\mathcal{L}_{\mathrm{dist}}$ を用いる．距離係数 $\alpha(d)$ を式 (2) に示す．

$$\alpha(d) = \alpha_{\mathrm{far}} + (\alpha_{\mathrm{near}} - \alpha_{\mathrm{far}}) \exp(-d/\tau) \tag{2}$$

ここで，d は自車両から検出物体までの距離である．また，α_{near}，α_{far} は近距離，遠距離に対する重みであり，τ は距離に対する減衰の強さを表す．さらに距離に対する重み付き損失関数 $\mathcal{L}_{\mathrm{dist}}$ を式 (3) に示す．

$$\mathcal{L}_{\mathrm{dist}} = \frac{1}{N} \sum_{j=1}^{N} \alpha(d_j)\, \mathcal{L}_j \tag{3}$$

ここで，$\mathcal{L}_j$ はサンプル j の検出損失，d_j はサンプル j までの距離，N はサンプル数を表す．式 (3) により，近距離ほど距離係数が大きくなり，損失への寄与が増加する．この距離に対する重み付き損失に基づき枝刈りを行い，評価値 $S(w_i)$ を式 (4) に示す．

$$S(w_i) = S^{\mathrm{dist}}_{\mathrm{class}}(w_i) + S^{\mathrm{dist}}_{\mathrm{SNIP}}(w_i) \tag{4}$$

ここで，$S^{\mathrm{dist}}_{\mathrm{class}} = \left|\frac{\partial \mathcal{L}_{\mathrm{dist}}(x^{\mathrm{specific}})}{\partial w_i} w_i\right|$ は距離重み付き特定クラス評価値，$S^{\mathrm{dist}}_{\mathrm{SNIP}} = \left|\frac{\partial \mathcal{L}_{\mathrm{dist}}}{\partial w_i} w_i\right|$ は距離重み付き SNIP 評価値である．式 (4) により，近距離物体の検出に寄与する重みが削除されず，近距離性能の劣化抑制が期待される．

4. 評価実験

BEVFormer に対して提案手法の有効性を検証する．定量評価では，枝刈り前後の精度を全距離/近距離 (0∼20 m) で比較する．定性評価では，検出結果の変化を確認する．

4.1. 実験条件

本実験では Backbone に ResNet101-DCN を用いる．学習設定はエポック数 24，バッチサイズ 64，枝刈り率 50%，70%，最適化は AdamW（lr=2e-4，weight decay=0.01）とし，データセットは NuScenes dataset を用いる．

4.2. 定量評価

全距離における枝刈りなしとありの精度を表 1，近距離における枝刈りなしとありの精度を表 2 に示す．表中の Methods は枝刈り時に距離係数を適用する距離帯（全距離／近距離）を表す．表 1 より，枝刈り率 70%では SNIP は重要な重みまで削除されるが，SNIP/ours(全距離) は特定クラスの重みを保持するため，SNIP および SNIP/ours(近距離) を上回った．また，表 2 より，枝刈り率 50%の近距離評価では，SNIP/ours(近距離) が近距離検出に重要な重みを保持するため，SNIP や SNIP/ours(全距離) を上回った．

表 1：全距離における定量評価

枝刈り率	Methods	NDS↑	mAP↑	mATE↓	mASE↓	mAOE↓	mAVE↓	mAAE↓
0%	枝刈りなし	0.524	0.417	0.655	0.273	0.367	0.354	0.194
50%	SNIP	**0.519**	0.413	**0.668**	0.276	**0.362**	**0.356**	0.207
50%	SNIP/ours(全距離)	0.516	**0.415**	0.698	**0.274**	0.384	0.375	**0.187**
50%	SNIP/ours(近距離)	0.513	0.410	0.703	0.275	0.395	0.390	**0.187**
70%	SNIP	0.491	0.386	0.720	0.284	0.493	0.448	0.198
70%	SNIP/ours(全距離)	**0.495**	**0.403**	**0.708**	**0.277**	**0.434**	**0.424**	**0.185**
70%	SNIP/ours(近距離)	0.487	0.381	0.736	**0.277**	0.450	0.470	0.198

表 2：近距離における定量評価

枝刈り率	Methods	NDS↑	mAP↑	mATE↓	mASE↓	mAOE↓	mAVE↓	mAAE↓
0%	枝刈りなし	0.608	0.562	0.564	0.267	0.338	0.343	0.218
50%	SNIP	0.605	0.552	0.583	**0.265**	0.327	0.322	**0.212**
50%	SNIP/ours(全距離)	0.595	0.537	0.589	0.270	0.396	0.330	0.223
50%	SNIP/ours(近距離)	**0.623**	**0.579**	**0.565**	0.267	**0.288**	**0.275**	0.270
70%	SNIP	0.594	0.541	**0.564**	**0.267**	0.369	**0.356**	**0.201**
70%	SNIP/ours(全距離)	0.596	0.546	0.657	0.274	0.348	0.357	0.203
70%	SNIP/ours(近距離)	**0.598**	**0.548**	0.605	**0.267**	**0.347**	0.359	**0.201**

4.3. 定性評価

枝刈り率 50%と枝刈り率 70%の可視化画像結果を図 2 と図 3 に示す．図 2 と図 3 より，SNIP/ours(近距離) の赤枠で示した歩行者の検出結果より，精度低下が抑制されていることが確認できる．

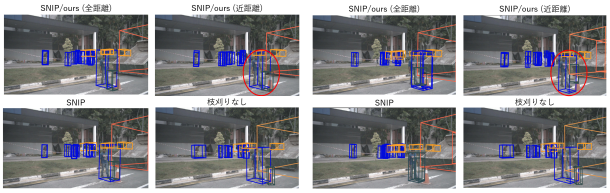


図 2：50%可視化画像

図 3：70%可視化画像

5. まとめ

本研究では，精度劣化を抑制した BEVFormer の重み付き非構造枝刈り手法を提案した．評価実験より，全距離評価では SNIP/ours(全距離) が SNIP を上回り，近距離評価では SNIP/ours(近距離) が NDS/mAP を改善した．今後は自動運転における評価として提案手法が運転経路に与える影響を検証する．

参考文献

[1] T. Ito *et al.*, “Weight Pruning to Mitigate Class-Specific Accuracy Degradation for LiDAR-Based 3D Object Detection,” in Proc. IEEE Intelligent Vehicles Symposium (IV), 2025.

[2] Z. Li, *et al.*, “BEVFormer: Learning Bird’s-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers,” In European Conference on Computer Vision (ECCV), 2022.

マスク誘導型 Visual Encoder による異常検知の言語説明と可視化

ER22075 森田悠斗　　指導教授：藤吉　弘亘

1. はじめに

工業製品の品質保証において，異常検知は重要工程である．異常検知は正常・異常の分類だけでなく異常領域の特定や判断根拠の可視化など，説明性も重要である．Anomaly-OneVision (Anomaly-OV) [1] は大規模言語モデル (LLM) を用いることで，異常内容を自然言語で説明することを可能とした．しかし，Anomaly-OV は異常内容の説明が異常領域と一致しないことがある．本研究では，異常領域を強調する学習を Visual Encoder に導入することで，異常内容の言語説明と異常領域の一致性を向上させる．

2. Anomaly-OV

Anomaly-OV は，入力画像から異常内容を自然言語で説明するマルチモーダルモデルである．本モデルは，Look-Twice Feature Matching (LTFM) と Visual Token Selector (VTS) を中核として構成される．LTFM は元画像と 4 分割して得られる局所画像の特徴を二段階で照合し，異常度スコアを算出する．VTS はそのスコアに基づき異常領域を強調した重み付き視覚特徴を生成し，LLM へ入力する．本モデルは，未知データの場合に異常を正確に捉えられず，説明文が異常内容から逸れる場合がある．

3. 提案手法

本研究では，教師マスクにより異常領域への注目を強化した Anomaly-OV に基づく異常説明と，その判断根拠の可視化手法を提案する．

3.1. 教師マスクによる Visual Encoder の拡張

Anomaly-OV の説明能力を特定の領域に適応させるため，欠陥箇所を明示したマスク画像を用いて追加学習を行う．提案手法のアーキテクチャを図 1 に示す．Visual Encoder には，元画像 I_0 およびそれを 4 分割した局所画像 I_n を入力する．Visual Encoder の出力である視覚特徴 v_j^0 を LTFM に入力し，異常度スコア m_j を出力する．そして，異常度スコア m_j と教師マスク y_j との誤差を最小化するよう LTFM を学習する．教師マスクは異常を 1，正常を 0 とする 2 値ラベルである．誤差関数には，Binary Cross Entropy (BCE) Loss を用いる．VTS では，異常度スコア m_j と視覚特徴 v_j^0 の要素積をクエリ，視覚特徴 v_j^0 をキー，バリューとして Q-Former で処理する．これを元画像 I_0 および局所画像 I_n に対して行う．異常領域を強調した特徴 v_j^s を画像ごとに LLM に入力し，異常説明文を生成する．

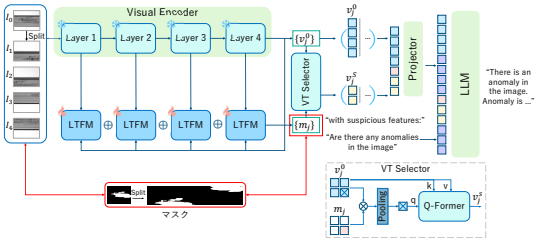


図 1: 提案手法のアーキテクチャ

3.2. 言語説明に対する可視化

異常説明文の生成時にモデルがどの画像および領域を参照しているか可視化する．各画像から得られる異常強調特徴の単語生成への寄与を比較するため，LLM の Attention weight を抽出し，注目画像を特定する．そして，異常度スコアに基づく異常強調特徴を生成する VTS の Q-Former における Cross-Attention から得られる注目度をパッチ単位で可視化することで，画像内の注目領域を特定する．

4. 評価実験

提案手法の有効性を示すために，異常ラベルの分類精度と説明性能の 2 つの観点で比較を行う．

4.1. 実験条件

評価には 10 種類の異常を含む電線データセットを用い，学習データ 8976 枚，評価データ 563 枚を使用する．学習条件は，学習率 $1e^{-4}$，バッチサイズ 16，10 エポックとする．また，異常説明生成に用いる LLM は Anomaly-Instruct125k でファインチューニングされたものを使用する．異常分類性能の評価では，従来のマルチモーダル手法である LLaVA，Anomaly-OV と提案手法を比較対象とする．評価指標として，生成された説明文と各異常カテゴリの正解文との一致度を用いる．生成文が正解文と 80%以上一致した場合を正解とみなし，正解率 (Accuracy) を算出した．異常説明性能の評価では，Anomaly-OV と提案手法を比較し，生成された異常説明文が異常特徴を適切に表現できているか定性的に評価する．

4.2. 異常分類結果

異常分類精度の比較結果を表 1 に示す．提案手法は LLaVA より 11.4 ポイント，Anomaly-OV より 0.5 ポイント高い精度を示した．従来手法では正常領域に注目することがあり，異常の識別が不十分となる場合があったが，提案手法は異常領域を強調することで誤分類が低減したと考えられる．

表 1: 異常分類精度の比較

	LLaVA	Anomaly-OV	提案手法
Accuracy	81.5	92.4	**92.9**

4.3. 異常説明性能の定性的評価

提案手法による説明文の質の向上について，異物_繊維の事例を用いて比較する．異物_繊維の出力結果を図 2 に示す．Anomaly-OV は，異物繊維の特徴を十分に反映できていない説明文である．一方，提案手法は,「波状」など繊維特有の異常形状を表す表現が生成され，異常特徴に基づく説明文が得られた．

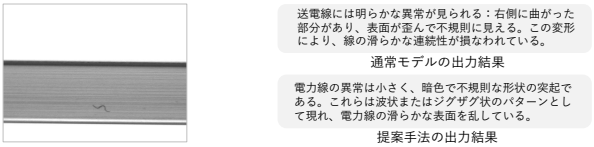
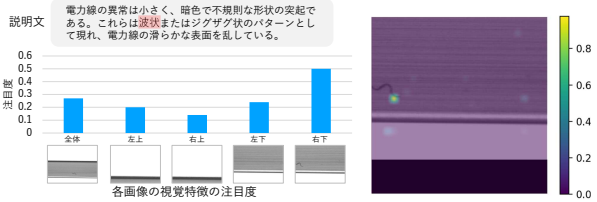


図 2: 異物_繊維の出力結果

4.4. 特定単語の注目領域の可視化

異常説明文内の「波状」という単語に着目し，単語生成時にモデルが参照する画像および領域を可視化した．注目画像および注目領域の可視化結果を図 3 に示す．図 3(a) より，欠陥箇所を含む右下の画像に注目度が集中しており，異常説明時に異常領域を多く含む画像を重視していることが分かる．図 3(b) の注目領域の可視化結果から，異常領域を含むパッチに注目が集中していることが分かる．これらの結果から，異常位置を明示的に学習させることで，モデルは異常領域に基づく単語生成を行うことを確認した．



(a) 注目画像の特定　　(b) 注目領域の可視化

図 3: 注目画像および注目領域の可視化結果

5. おわりに

本研究では，Anomaly-OV を電線データに適用し，異常領域の強調と説明文生成の改善を行った．異常度スコアに対してマスク画像を用いた追加学習を実施し，モデルは異常領域をより正確に捉え，説明文も適切な異常表現へと改善された．今後は，異常領域の推定精度向上や，単語生成時の注目領域をより詳細に可視化する手法の検討を進める．

参考文献

[1] Y. Xu, *et al.*, “Anomaly-OV: Towards Zero-Shot Anomaly Detection and Reasoning with Multimodal Large Language Models”, In CVPR, 2025

学習行動の文脈を考慮した教師なし異常行動検知

ER22076 安田恭大　　指導教授：藤吉弘亘

1. はじめに

教材のデジタル化に伴い，電子教材上での学生の学習行動ログを大規模に収集することが可能となった．この学習行動ログデータを分析することで，学生一人一人に合わせた学習サポートの実現が期待されている．先行研究では，行動ごとの出現回数をヒストグラム特徴とした各学生の成績予測手法が提案されている．成績予測は最終的な成績を予測するものであり，学習行動の順序に起因する異常行動や特異性を捉えることは困難である．そこで本研究では，多くの学生が行う標準的な学習行動から逸脱した行動系列を異常と定義し，学習行動の文脈情報を捉える埋め込み表現による教師なし異常検知手法を提案する．

2. 先行研究

小濱らは，電子教材における行動ごとの出現回数をヒストグラム特徴として表現して成績予測を行う手法を提案した [1]．この手法は成績予測に有効である一方，行動の出現回数に依存した予測のため，学習行動における行動の違いや文脈的特異性を捉えることが難しく，学習プロセスの分析には課題がある．

3. 提案手法

本研究では，学習行動における文脈的な異常を捉えるため，行動系列の文脈情報を埋め込みベクトル化し，Isolation Forest[2] を用いて異常行動を検知する手法を提案する．本手法は，Masked Language Model (MLM) によるモデル学習，ベクトル生成，異常検知の 3 段階で構成される．提案手法の概要図を図 1 に示す．

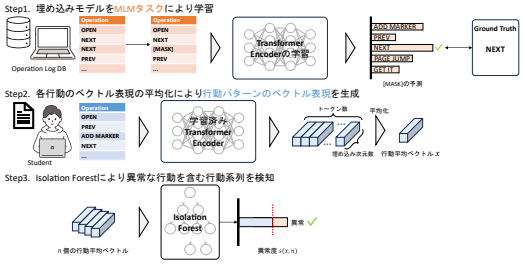


図 1: 提案手法の概要図

3.1. 行動系列の埋め込みベクトル生成

Step1 では，系列内の離れた行動間の関係性を捉えるため Transformer Encoder を採用し，一部をマスクした行動系列を入力として MLM により学習を行う．Step2 では，学習したモデルに行動系列を入力し，各行動に対応する文脈情報を保持した埋め込みベクトルを得る．このベクトルを平均化することで，行動平均ベクトルを生成する．

3.2. Isolation Forest による異常行動の検知

Step3 では，生成したベクトルを Isolation Forest に入力し，正解ラベルを用いずに特徴空間上で孤立した異常な行動パターンを持つ行動系列を検知する．Isolation Forest は，決定木に基づく教師なし異常検知手法である．学習時は，ランダムに選択された特徴量と，その最大値・最小値の間からランダムに決定された分割点を用いてデータを再帰的に二分割し，多数の決定木を構築する．各データが葉ノードに孤立するまでの平均パス長に基づき，式 (1) で異常度を算出する．

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

ここで，$h(x)$ はデータ x が葉ノードに到達するまでのパス長，$E(h(x))$ はその平均値，$c(n)$ はデータ数 n における平均パス長である．正常データは孤立までに多くの分割を要する一方，異常データは少ない分割で孤立するためパス長が短くなる傾向がある．この性質を利用して異常行動を検知する．

4. 評価実験

本実験では，提案手法を用いて異常行動パターンを検知する．また，検知したデータの行動履歴を用いてその行動の分析を行う．

4.1. 実験条件

本実験では，九州大学で収集された学習行動ログデータセットを用いる．訓練データに 2019 年から 2021 年（1,209 名）までのデータを，評価データに 2022 年（237 名）のデータを使用し，いずれも 1 から 8 週目までの講義時間内のデータに限定して実験を行う．モデルは，BERT-Base の構成（入力 512 トークン，次元数 768，12 層）を基にした Transformer Encoder を使用する．学習には，学習率を 1e-5，バッチサイズを 32，エポック数を 300 に設定した．損失関数には，各行動の希少性に基づく重みである IDF を適用した Weighted Cross Entropy Loss を用いる．また，異常判定の閾値は，異常度の上位 5%に設定する．

4.2. 実験結果

異常検知の結果を表 1 に示す．表 1 より，全データ数 1,463 件に対し，74 件が異常と判定された．正常データの平均行動回数は約 152.2 回であるのに対し，異常データは平均 30.6 回と，系列長が短い傾向が見られた．また，成績分布を確認すると，異常データ群では成績 F や D の割合が高い傾向が見られる．一方，成績 A の学生も異常データの約 3 割（28.4%）を占めている．これは，標準的な学習行動とは異なる特異な行動パターンを持つ週や，遅刻や早退によって学習時間が極端に短くなり，行動回数が少なくなった週が検知されたためと考えられる．

表 1: 異常度（平均パス長）に基づく検知結果

判定結果	データ数	割合 [%]	平均行動回数	成績分布 [%]				
				A	B	C	D	F
正常	1,389	94.9	152.2	38.7	27.1	19.8	9.8	4.6
異常	74	5.1	30.6	28.4	16.2	23.0	21.6	10.8

4.3. 定性的評価

行動平均ベクトルの異常度の分布を図 2 に，各クラスタを色で表示した結果を図 3 に示す．図 2 より，異常度が高いデータが空間の左端に密集し，独立した領域を形成していることがわかる．この領域を図 3 のクラスタリング結果と照合すると，行動回数が少なく，“OPEN”（教材を開く）や “CLOSE”（教材を閉じる）といった行動が主体であるクラスタに対応している．これは，システムへのアクセス後に内容を読み進めることなく即座に離脱した行動を示しており，多くの学生に見られるページ遷移の文脈を持たないため，特徴空間上で孤立し異常と判定されたと考えられる．一方，正常データに対応するクラスタは，行動回数が多く，“NEXT”（次のページに移動）を主体としたページ遷移を多く伴う行動系列であることが確認された．

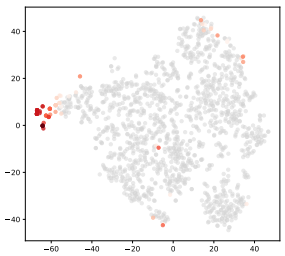


図 2: 異常度の空間分布

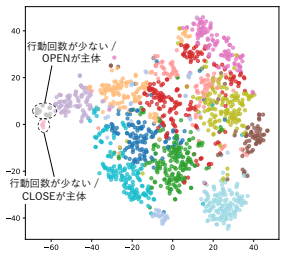


図 3: クラスタ分布

5. おわりに

本研究では，文脈を考慮した行動特徴と Isolation Forest を用いた異常行動の検知および分析を行った．実験の結果，異常データは即座離脱による文脈欠如が多く，成績 F・D の割合が高かったのに対し，正常データはページ遷移を多く伴い，成績 A の割合が高かった．このことから，ページ遷移を伴う継続的な学習行動が成績向上に寄与すると考えられる．今後は，ページ滞在時間などの時間情報を加え，閲覧行動の質的な差異を考慮した異常検知を目指す．

参考文献

[1] H. Kohama, *et al.*, “Recommending Learning Actions Using Neural Network”, ICCE, 2023.

[2] F. Liu, *et al.*, “Isolation Forest”, ICDM, 2008.