



# スポーツ動作の差分測定と ユーモア動画生成への応用

植木研究室  
21J5-083 田島 友希

スポーツ動作において、自分の**特徴**を見つけることは難しい。

骨格推定技術を使用して**定量的に評価**することができる。

スポーツ動作解析技術を**応用した研究**が少ない。

スポーツ動作×ユーモアの**新しい分野**に応用する。



モノマネ芸人



動きが面白い?  
雰囲気面白い?  
見た目が面白い?

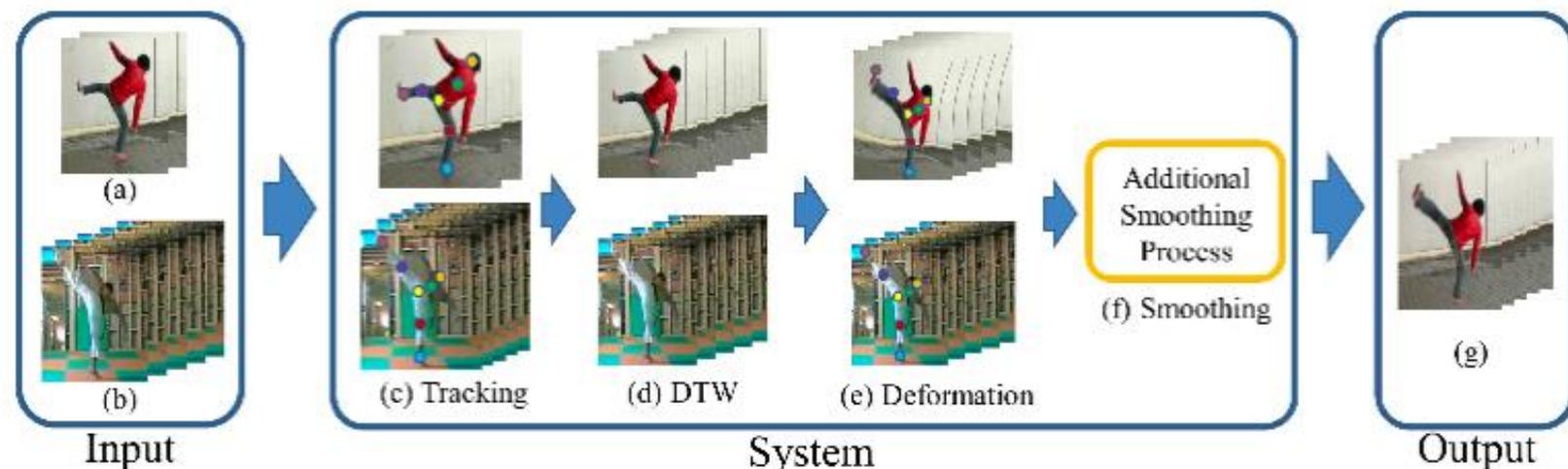
# “Choreographing Amateur Performers Based on Motion Transfer between Videos”

Kenta Mizui, Makoto Okabe, Rikio Onai

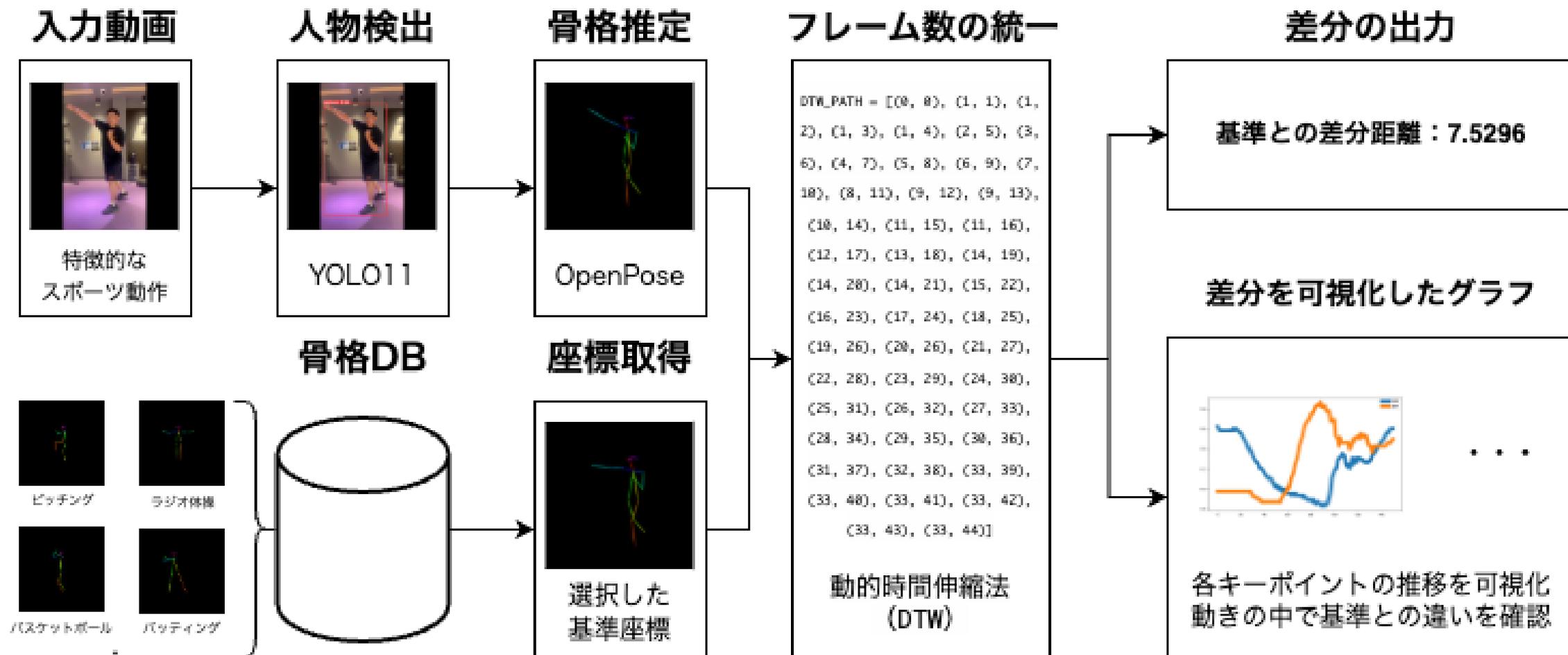
International Conference on Advances in Multimedia Modeling (MMM 2013)

キーフレーム抽出とその空間的・時間的な改変を通じて、動きのある重要なシーンを誇張し、ビデオの要点を効果的に表現する。

実験結果では、プロの動画と視覚的に非常に近い動きを再現できることが示されている。



# 差分測定システム概要



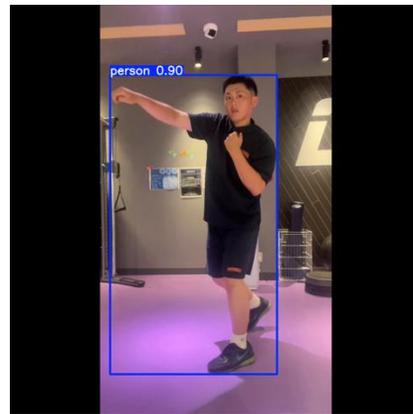
## 人物検出

## 骨格推定



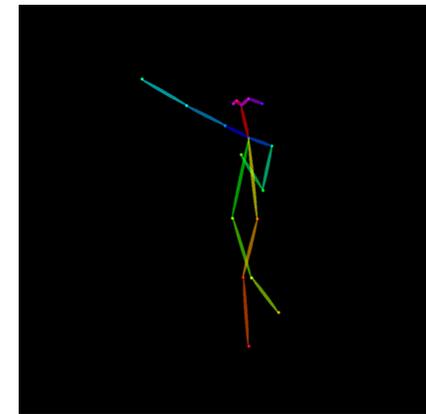
## 入力画像

ボクシングのパンチを  
している人物画像



## YOLO11

人物を検出  
領域の座標を取得する



## OpenPose

目, 鼻, 左右の腕といった  
18部位のキーポイントを検出

## フレーム数の統一

## 動的時間伸縮法 (DTW)

2つの動画の

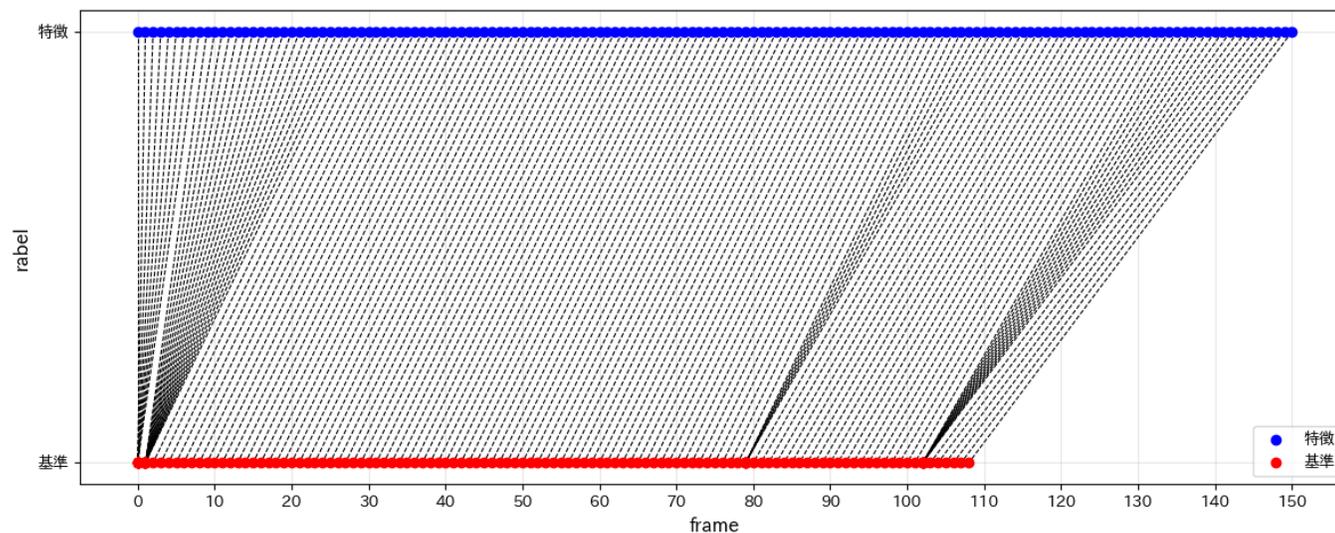
キーポイント座標を DTW を使用して

組み合わせを決定し,

動画間における全フレームの

キーポイント同士の距離が

最短となるフレーム数に調整する。



# 検証実験

システムの有効性を検証するために実験を行った。

動画1をお手本とし、各動画とのキーポイントの距離を計測することで動作がどれだけ類似しているかを検証する。

入力動画：画面サイズ：640×640 ・フレームレート：60

種類	内容	フレーム数
動画1	野球のスイング（お手本）	204
動画2	動画Aの模倣スイング	156
動画3	野球の異なるスイング	178
動画4	大きなジャンプ動作	170



動画1



動画2



動画3



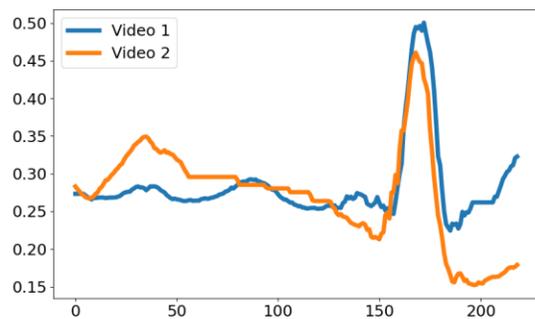
動画4

# 実験結果

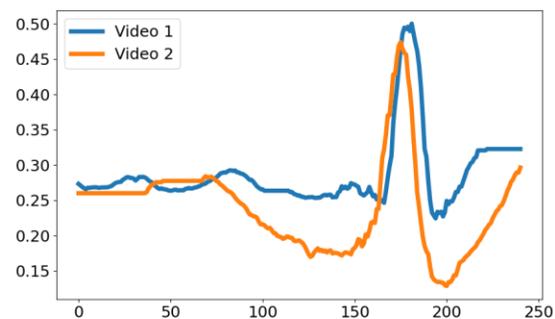
動画1と比較した動画	キーポイント間距離 (全体)
動画2	7.0569
動画3	7.8734
動画4	8.4140



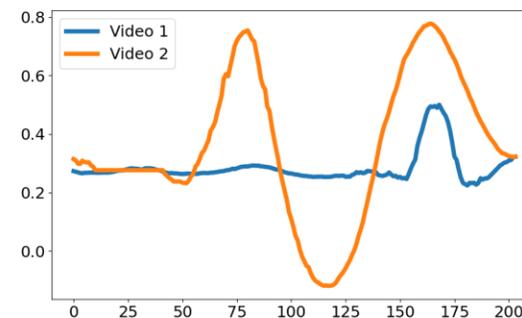
動画1



動画1&動画2

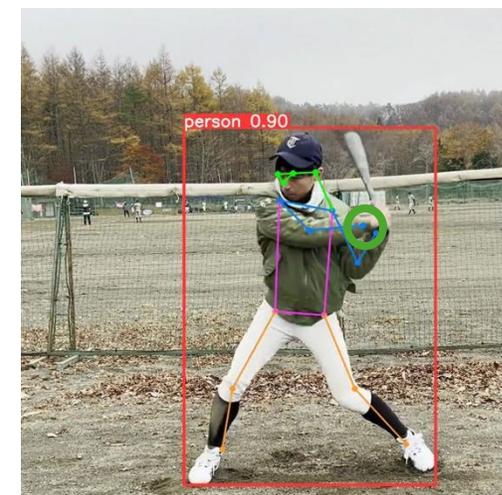


動画1&動画3



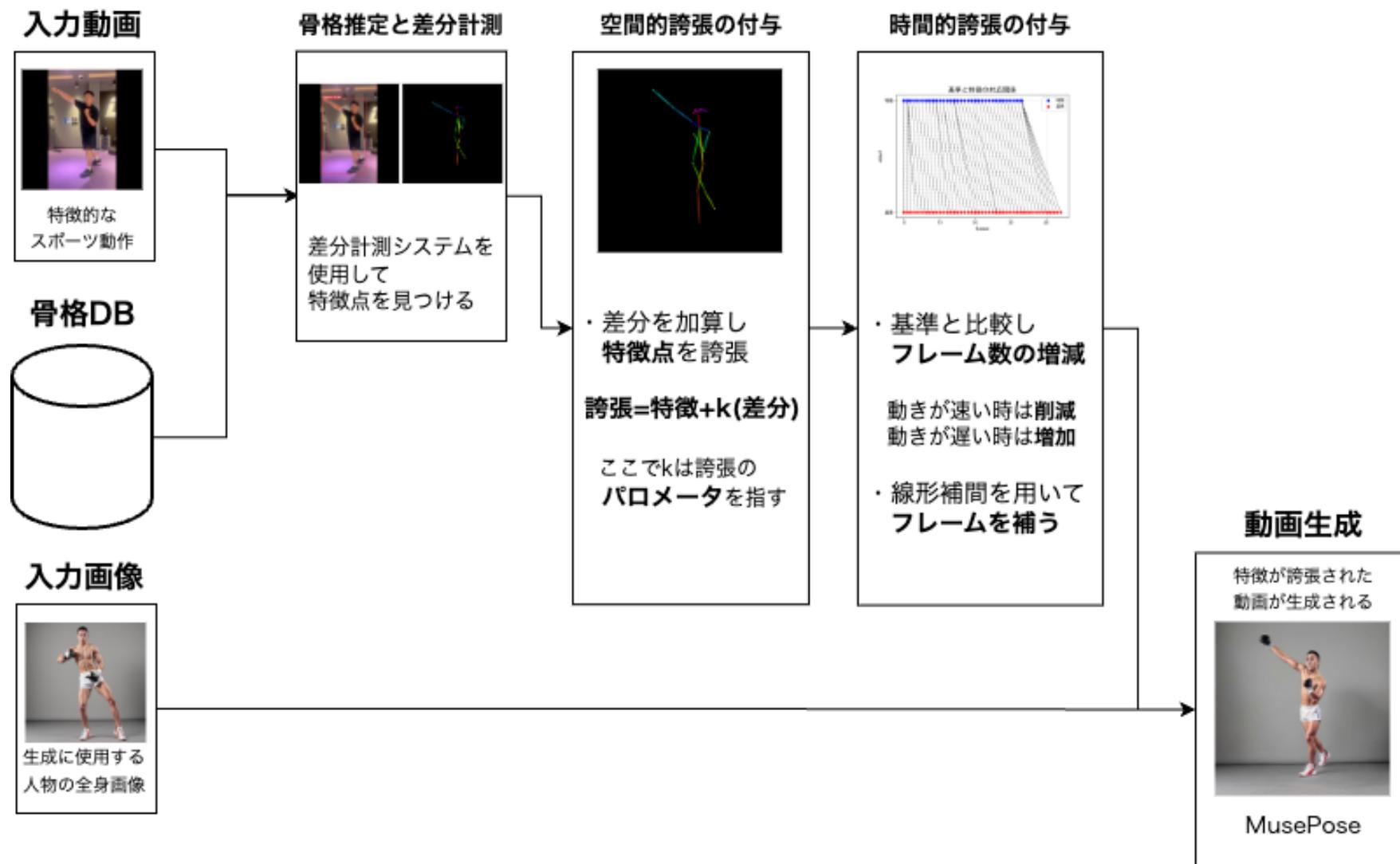
動画1&動画4

左手首のY軸の距離を示したグラフ



動画2 8

# ユーモア動画生成システム概要



# 検証実験

## ・入力動画

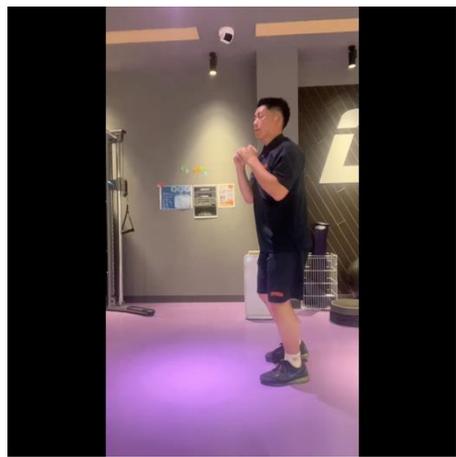
種類	スポーツ動作	特徴
動画1	バスケットボールのフリースロー動作	深くゆっくりしゃがみ込んでスローする
動画2	ボクシングのパンチ動作	素早く高くパンチする
動画3	ラジオ体操の腕を振って足を曲げ伸ばす運動	腕を高く上げる
動画4	野球のプロアスリートの打撃動作を模倣した動作	ゆっくり足を高く上げ前に踏み込む
動画5	野球のプロアスリートの投球動作*	足を高く上げて2段階で前に踏み込む

動画名から動画を見ることができます

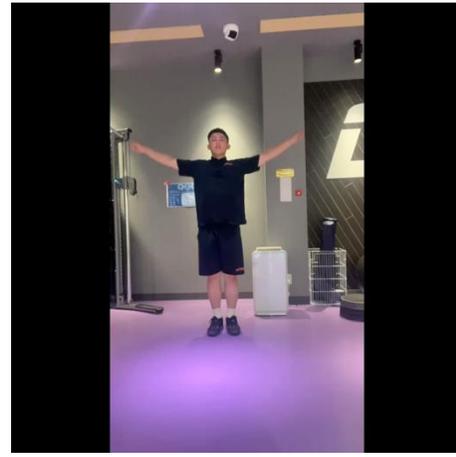
\*<https://www.youtube.com/watch?v=NDFr8veGcv8>



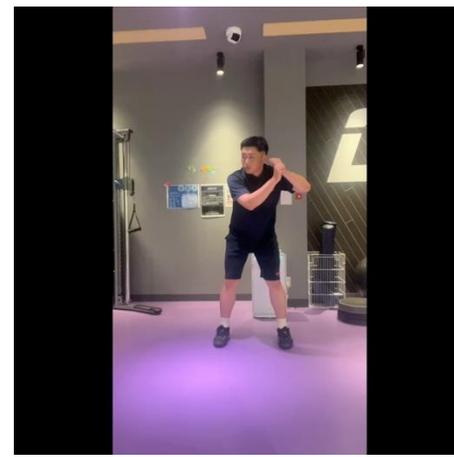
動画1



動画2



動画3



動画4



動画5

# バスケットボールのフリースロー動作 結果



入力画像

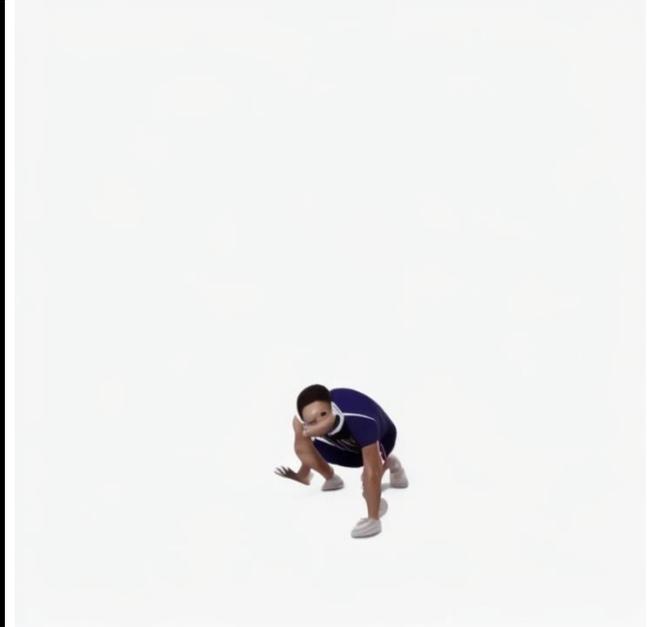
誇張された生成動画

- 空間的特徴
- 時間的特徴
- 時空間的特徴

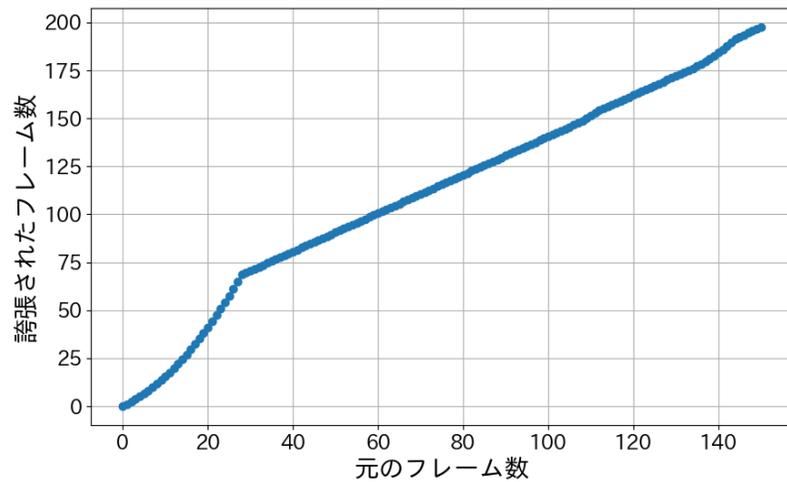
動画名から動画を見ることができます



元動画



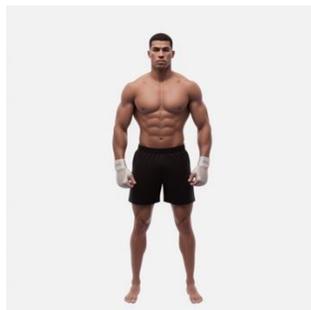
生成動画



フレーム数増加の推移

特に誇張されたフレームの比較

# ボクシングのパンチ動作 結果



入力画像

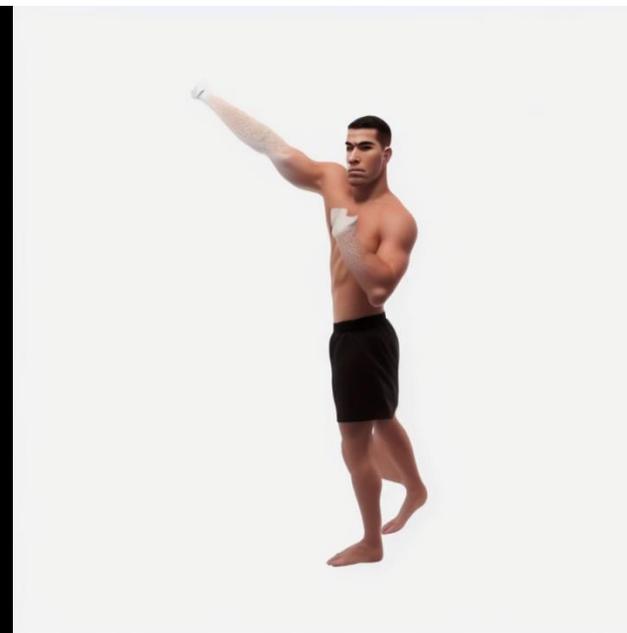
誇張された生成動画

- 空間的特徴
- 時間的特徴
- 時空間的特徴

動画名から動画を見ることができます

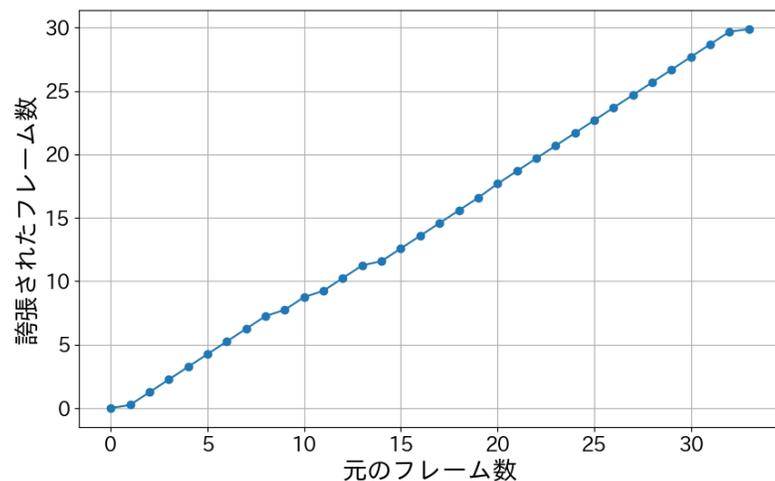


元動画



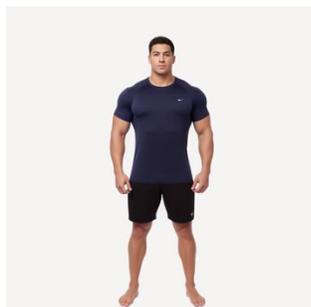
生成動画

特に誇張されたフレームの比較



フレーム数増加の推移

# ラジオ体操の腕を振って足を曲げ伸ばす運動 結果

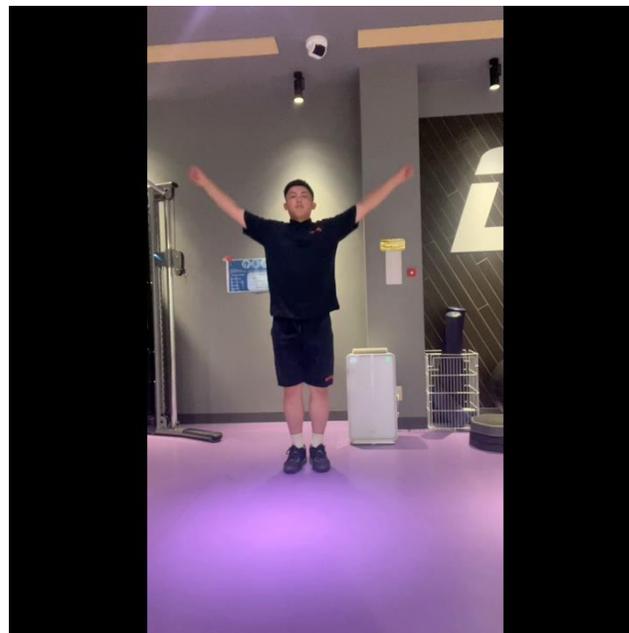


入力画像

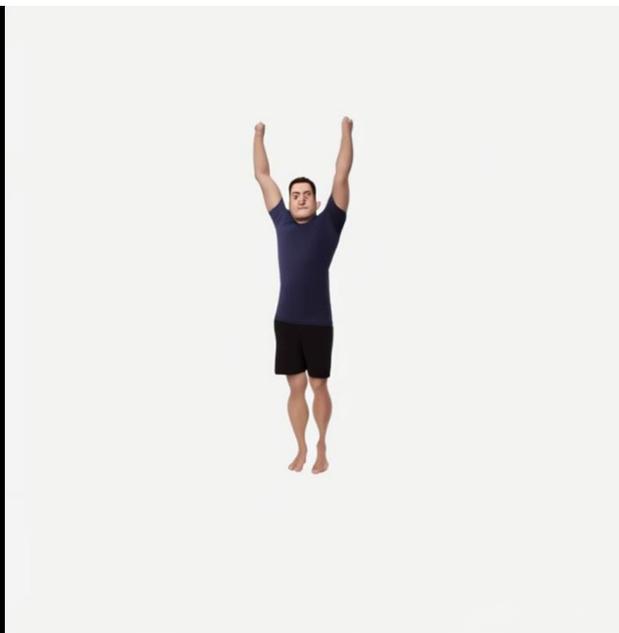
誇張された生成動画

- 空間的特徴
- 時間的特徴
- 時空間的特徴

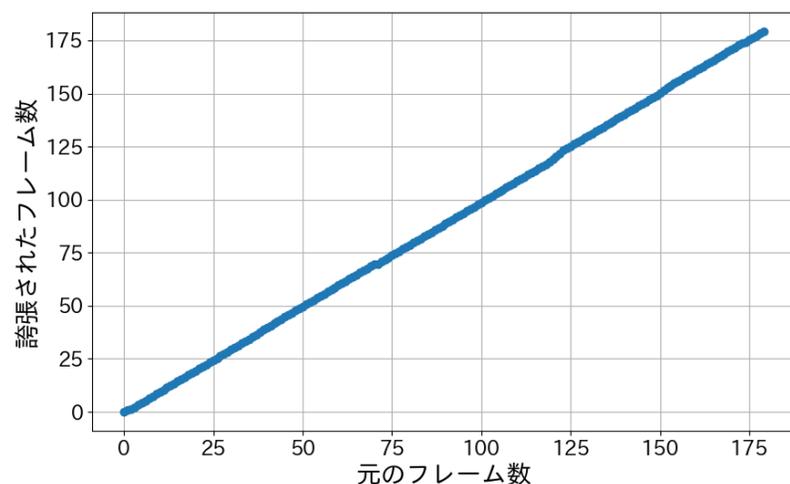
動画名から動画を見ることができます



元動画



生成動画



フレーム数増加の推移

特に誇張されたフレームの比較

# 野球のバッティング動作 結果



入力画像

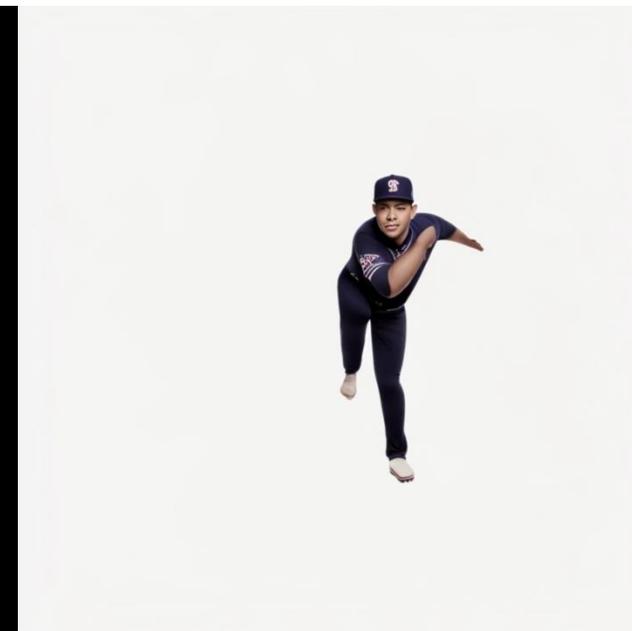
誇張された生成動画

- 空間的特徴
- 時間的特徴
- 時空間的特徴

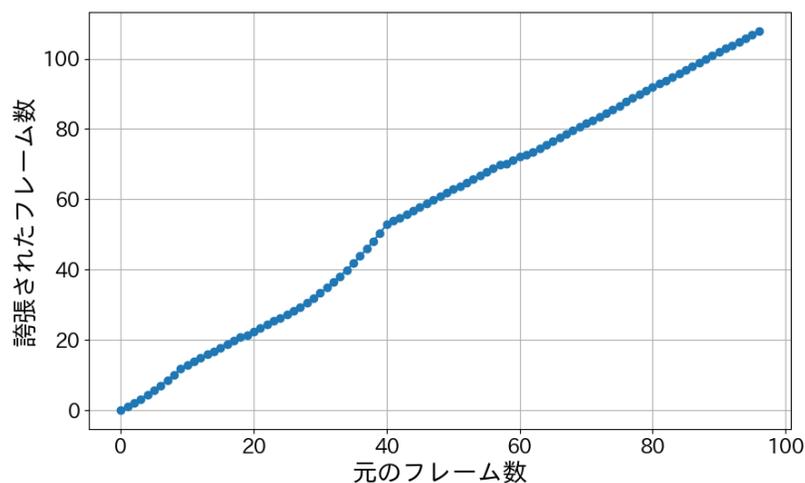
動画名から動画を見ることができます



元動画



生成動画



フレーム数増加の推移

特に誇張されたフレームの比較

# 野球のピッチング動作 結果



入力画像

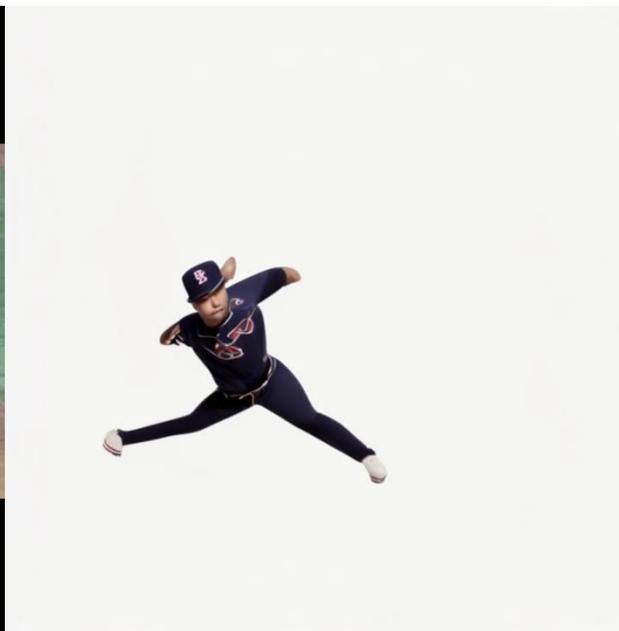
誇張された生成動画

- 空間的特徴
- 時間的特徴
- 時空間的特徴

動画名から動画を見ることができます

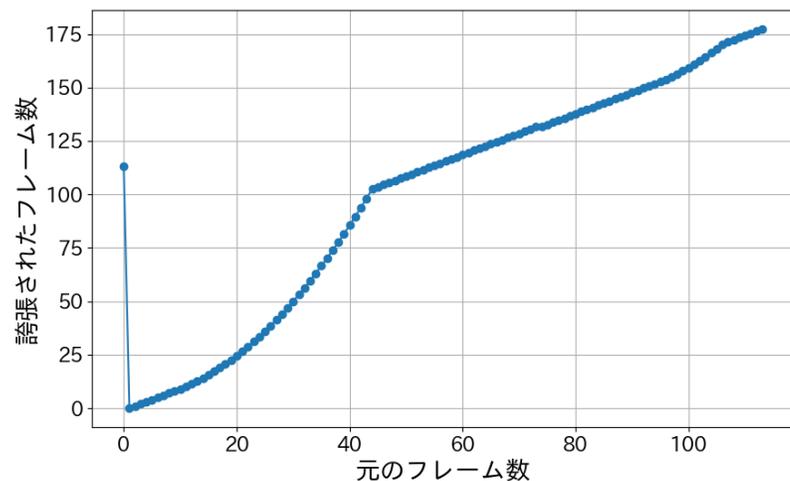


元動画



生成動画

特に誇張されたフレームの比較



フレーム数増加の推移

# 生成動画に対する評価アンケート

日時：2024年12月12日～2024年12月25日

年代：10代～20代

人数：52人

## 評価方法

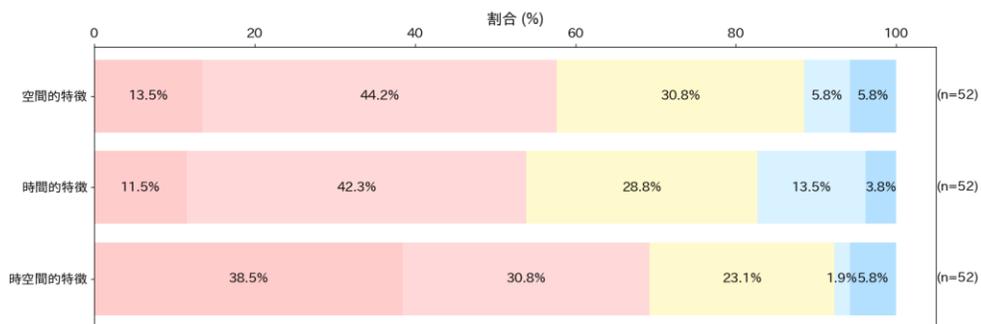
元の特徴動画を見た後、各誇張を与えた動画を見て五段階で評価をする。

「ユーモアである」「ややユーモアである」

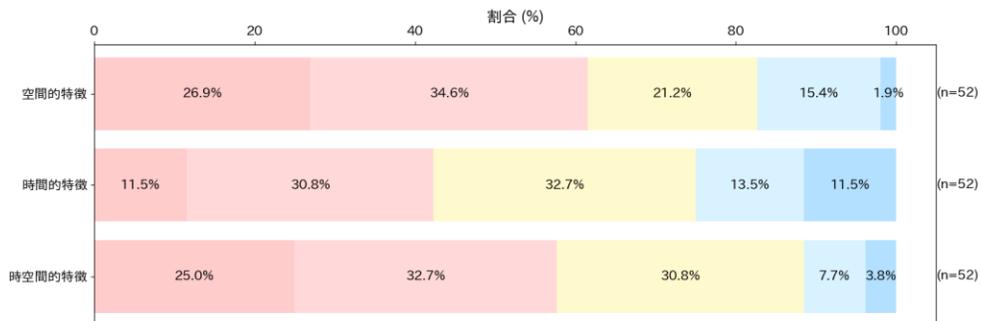
「どちらでもない」

「ややユーモアではない」「ユーモアではない」

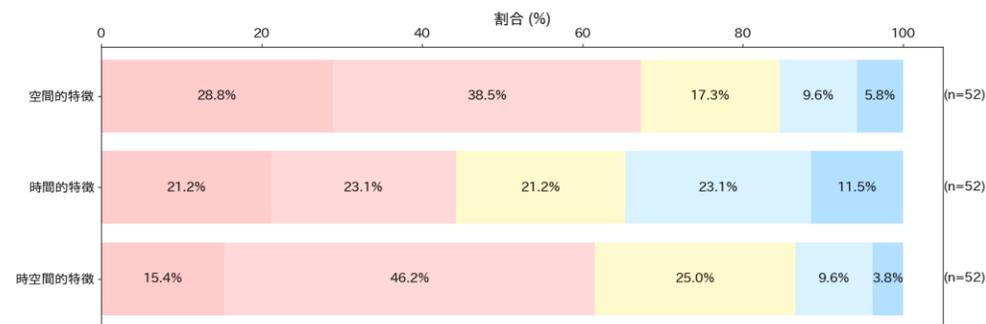
# アンケート結果



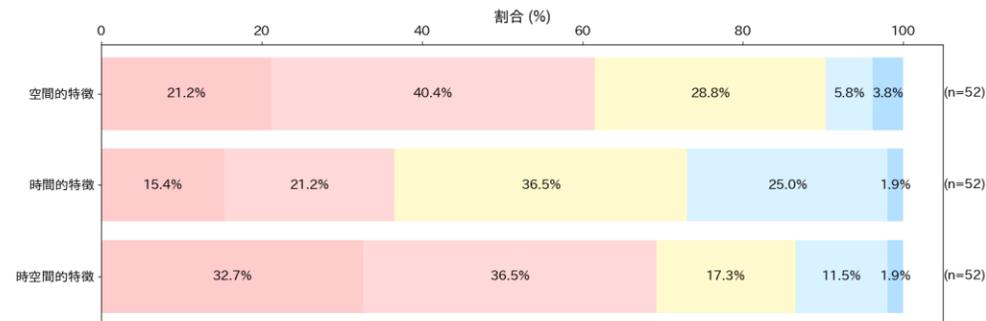
## フリースロー動作



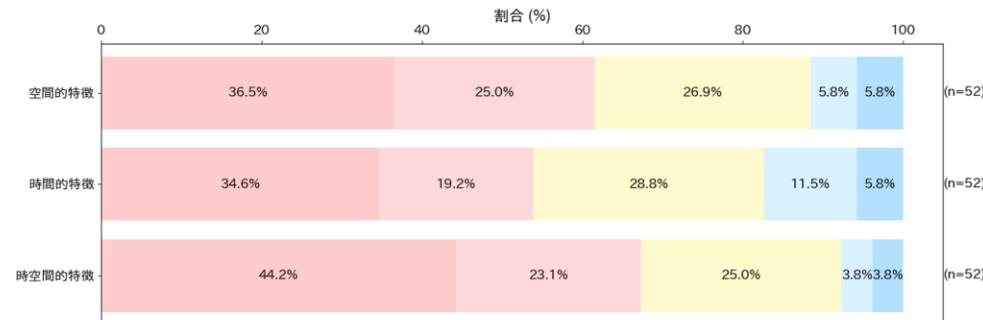
## ラジオ体操



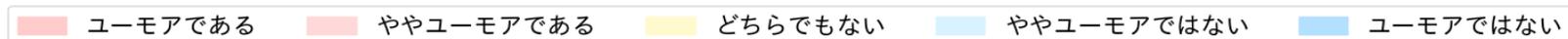
## パンチ動作



## バッティング動作



## ピッチング動作



# 考察

- 空間的特徴  
全ての動作において**ユーモアに作用した**.
- 時間的特徴  
**動画1と動画5**には**ユーモアに作用した**.  
**動画2, 動画3, 動画4**では**ユーモアにあまり作用しない**.
- 時空間的特徴  
全ての動作において**ユーモアに作用するした**.

**空間的特徴, 時空間的特徴**→**ユーモアに作用**

**時間的特徴**→**単独ではユーモアには限定的**

## まとめ

### 差分測定手法の提案

動的時間伸縮法（DTW）により、異なるフレーム数の動画間で**動作の類似性**を高精度に評価可能。

### ユーモア動画生成システムの開発

空間・時間的特徴の**誇張**による**ユーモア性の高い動画**を生成し、評価実験で**高評価**を獲得。

---

## 今後の課題

### システムの改良化

**生成品質の向上**や**リアルタイム処理**の実現。

### ユーモア性評価の客観化

**幅広い年代**と**多人数**へのアンケートを実施する。

## 発表実績

「スポーツ動作の差分測定システム」

動的画像処理実利用化ワークショップ(DIA2024), 2024.

**田島 友希**, 武藤 良, 植木 一也

「スポーツ動作の骨格推定とユーモア動画生成への応用」

画像の認識・理解シンポジウム(MIRU2024), 2024.

**田島 友希**, 松山 諒大, 武藤 良, 植木 一也

「スポーツ動作の特徴誇張システム」

芸術科学会(NICOGRAPH2024), 2024.

**田島 友希**, 小松 祐也, 武藤 良, 植木 一也

「多様なテキスト・画像モデルを活用した文字アートの自動生成」

芸術科学会(NICOGRAPH2024), 2024.

日高 健士郎, **田島 友希**, 松山 諒大, 武藤 良, 植木 一也

「SMoEx(エスモエクス) (Sports Motion Exaggeration) : スポーツ動作の特徴誇張動画の生成」

動的画像処理実利用化ワークショップ(DIA2025), 2025. (発表予定)

**田島 友希**, 武藤 良, 植木 一也



**ご清聴ありがとうございました**

Thank you for listening.

## 参考資料

# WebシステムSMoEXの構築



100フレームほどの動画であれば6分ほどで生成可能。  
最大VRAM使用量は31GB.

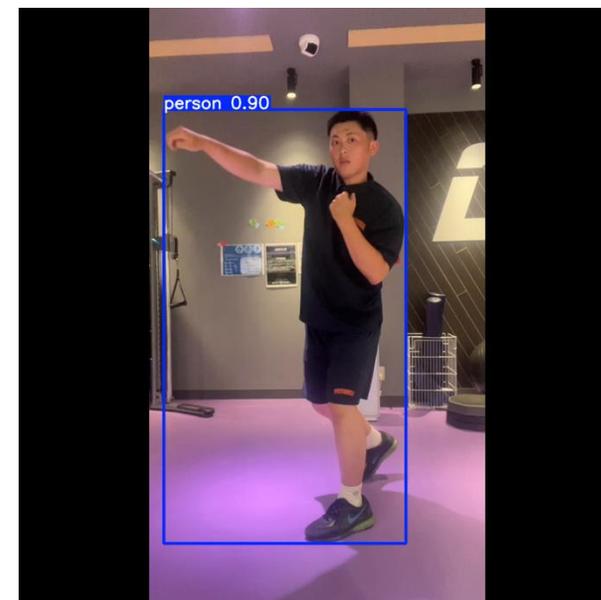
## 参考資料

人物検出：[YOLO11](#)

高速かつ高精度な物体検出を実現する、最新のディープラーニングモデル。



検出



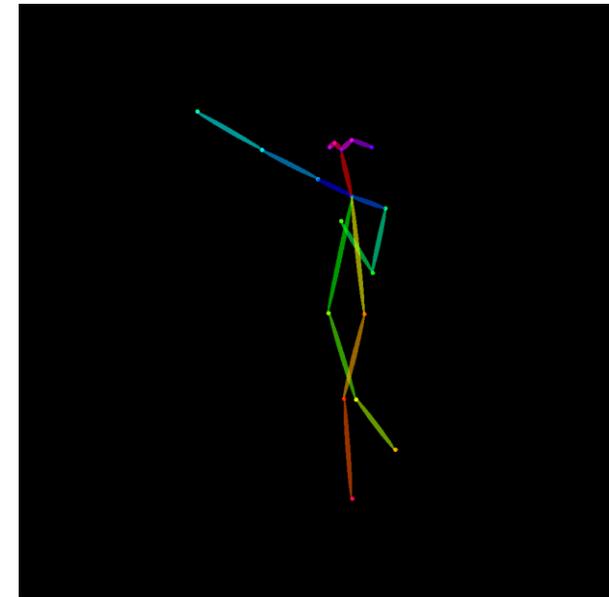
## 参考資料

骨格推定：OpenPose

画像中の人物に対し，目，鼻，左右の腕といった18部位のキーポイントを検出する。



検出



## 参考資料

# 空間的特徴

## 差分測定

$$\Delta x = x_1 - x_2, \quad \Delta y = y_1 - y_2$$

### 内訳:

- $x_1, y_1$  特徴的な動作の座標
- $x_2, y_2$  標準的な動作の座標

**意図**：動作の違いを数値化し，誇張の基準にする。

## 参考資料

# 空間的特徴

## 差分測定

$$x'_1 = x_1 + k \cdot \Delta x, \quad y'_1 = y_1 + k \cdot \Delta y$$

### 内訳:

- $k$ は誇張倍率
- $x'_1, y'_1$ は誇張後のキーポイント座標

**意図**：差分を基準に，動画の特徴を指定した倍率で誇張。

## 参考資料

# 時間的特徴

## フレーム短縮 (Shortening)

$$F_{short}[i] = F_{short}[i - 1] + \frac{1.0}{C_s}$$

### 内訳:

- $i$ は現在処理中のフレームインデックス ( $i \geq 1$ )
- $F_{short}[i - 1]$ は直前の短縮後のフレーム位置
- $C_s$ は短縮対象のフレーム総数

**意図** : 入力動画のフレームを圧縮し、**視覚的に**入力動画の**速度を上げる**

## 参考資料

# 時間的特徴

## フレーム延長を行う (Extending)

$$F_{ext}[i] = F_{orig}[i] + \frac{1.0}{C_e}$$

### 内訳:

- $F_{orig}[i]$ は変換前のフレーム位置
- $F_{ext}[i]$ は変換後のフレーム位置
- $C_e$ は延長対象のフレーム位置

**意図**：フレーム間隔を均等に広げることが可能

## 参考資料

## 時間的特徴

**線形補間：2点間を直線で結び、その間の任意の点を計算する補間手法**

新しいフレーム位置を $F_{new}$ とし、挿入対象となる両端のフレーム位置を $F_{low}$ （低い側）および $F_{high}$ （高い側）とする。

$$F_{new} = R_b \cdot F_{low} + R_a \cdot F_{high}$$

- $R_b$ は低い側のフレーム距離比率
- $R_a$ は高い側のフレーム距離比率

各比率の計算

$$R_b = \frac{D_{low}}{D_{low} + D_{high}}, \quad R_a = \frac{D_{high}}{D_{low} + D_{high}}$$

- $D_{low}$ は新たなフレーム位置と低い側のフレーム位置との距離
- $D_{high}$ は新たなフレーム位置と高い側のフレーム位置との距離

## 参考資料

## 時空間的特徴

処理の統合：

空間的誇張：動作の位置的特徴を強調。

時間的誇張：動作速度やテンポを調整。

例:空間的誇張：ポーズ画像の特徴を強調。

時間的誇張：動作の速度変化や自然な遷移を表現。

DTW（動的時間伸縮法）：

フレームの対応関係を利用して、短縮・延長・補間を統一的に実施。

## MusePose: 高品質動画生成システム

### 概要 :

- 動作の特徴と外観情報を統合し, リアルな動画を生成.
- 時間的一貫性とスタイル整合性を確保.

VAE : データの圧縮と生成。

2D/3D Unet : 空間・時間的特徴を抽出。

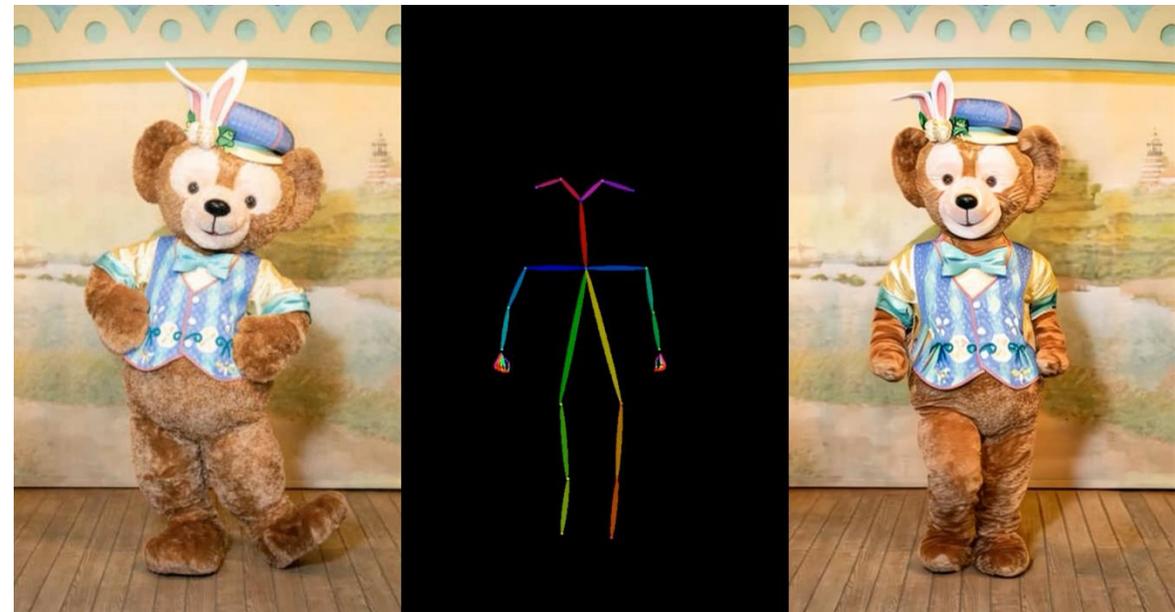
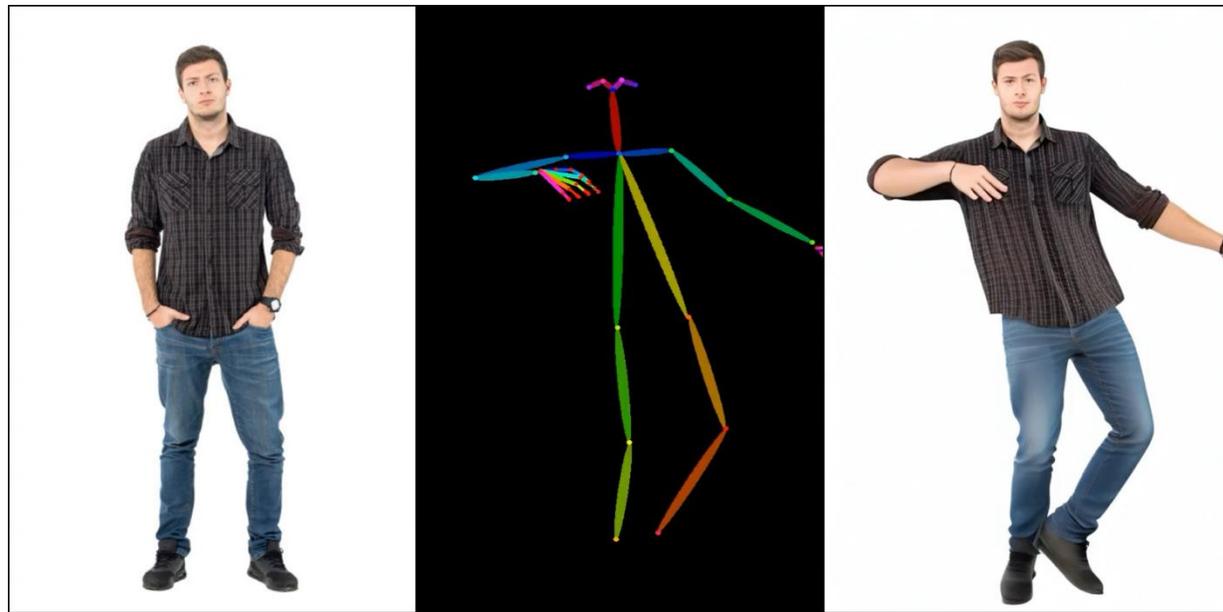
CLIP Vision Encoder : 視覚情報とテキストの意味一致を実現。

Temporal Attention : 滑らかな動きの生成。

Style Transfer Module : 背景と前景の統一感を維持。

## 参考資料

# MusePose: 高品質動画生成システム



引用：[MusePoseのgitリポジトリ](#)

## 参考資料

## 評価優位性を求めるための検定

- ポジティブ評価
  - ユーモアである
  - ややユーモアである
- 中立評価
  - どちらでもない
- ネガティブ評価
  - ユーモアではない
  - ややユーモアではない

各カテゴリの分布を比較評価  
カテゴリ間の有意差を検証

## 参考資料

## $\chi^2$ 検定

目的：評価カテゴリの分布差の有意性検定

仮説：

- 帰無仮説 ( $H_0$ )：分布に有意差はない
- 対立仮説 ( $H_1$ )：分布に有意差がある

結果：検定結果に基づき次の手法で詳細比較

## 参考資料

## 多重比較（ライアンの名義水準法）

目的：評価カテゴリ間の詳細な差異を検証

仮説：

- 帰無仮説 ( $H_0$ )：カテゴリ間に有意な差はない
- 対立仮説 ( $H_1$ )：カテゴリ間に有意な差がある

結果：有意水準  $\alpha = 0.05$ （両側検定）

## 参考資料

## 動作1 (バスケットボールフリースロー)

結果 :

- 半数以上が「ユーモアである」「ややユーモアである」と回答
- $\chi^2$ 検定：すべての誇張動作で有意な差あり ( $p < 0.01$ )

多重比較：

- 空間的特徴、時間的特徴、時空間的特徴でそれぞれ有意差あり
- 時空間的特徴では複数の評価カテゴリ間に有意差

## 参考資料

## 動作2 (ボクシングパンチ動作)

結果 :

- 空間的特徴、時空間的特徴で「ユーモアである」が半数以上
- 時間的特徴では半数以下

 $\chi^2$ 検定 :

- 空間的特徴、時空間的特徴で有意な差あり ( $p < 0.01$ )
- 時空間的特徴では複数の評価カテゴリ間に有意差 ( $p > 0.05$ )

多重比較 :

空間的特徴と時空間的特徴で詳細な有意差

## 参考資料

## 動作3（ラジオ体操動作）

結果 :

- 空間的特徴、時空間的特徴で半数以上がポジティブ評価
- 時間的特徴では半数以下

$\chi^2$ 検定 :

- 空間的特徴、時空間的特徴で有意な差あり ( $p < 0.01$ )
- 時空間的特徴では複数の評価カテゴリ間に有意差 ( $p > 0.05$ )

多重比較 :

空間的特徴がユーモア認識に大きく寄与

## 参考資料

## 動作4 (バッティング動作)

結果 :

- 空間的特徴、時空間的特徴で半数以上がポジティブ評価
- 時間的特徴では半数以下

$\chi^2$ 検定 :

- 空間的特徴、時空間的特徴で有意な差あり ( $p < 0.01$ )
- 時空間的特徴では複数の評価カテゴリ間に有意差 ( $p > 0.05$ )

多重比較 :

空間的特徴と時空間的特徴で詳細な差異を確認

## 参考資料

## 動作5（ピッチング動作）

結果：

- 半数以上が「ユーモアである」「ややユーモアである」と回答
- $\chi^2$ 検定：すべての誇張動作で有意な差あり ( $p < 0.01$ )

多重比較：

- 空間的特徴、時間的特徴、時空間的特徴でそれぞれ有意差あり
- 時空間的特徴では複数の評価カテゴリ間に有意差

## 参考資料

## アンケート行った上での自由記述

## ポジティブな意見

- 動作によって大きくユーモアが変わっていて面白いと感じました。
- さらに色々なスポーツ動作を見てみたい

## ネガティブな意見

- 誇張された動画が乱れていてよく分からない映像に感じる。



全ての動作においてユーモアではないと評価された