

# 深層学習に基づく行動に着目したシーン抽出手法

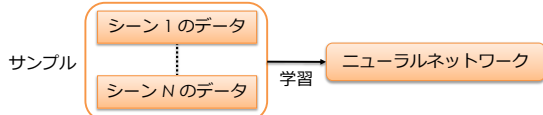
山西 宏平

大阪大学基礎工学部情報科学科

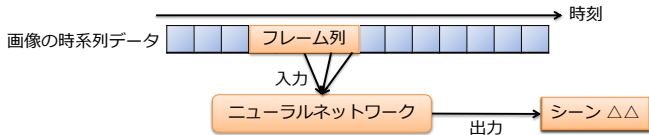
村田研究室

## 研究目的

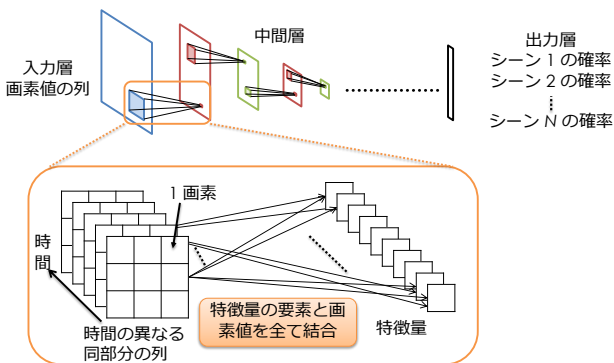
- 深層学習を「動画データから各時刻のシーンを抽出・識別するアプリケーション」に応用
  - 各シーンごとに幾つかのサンプルを用意し、オフラインで学習



- オンラインで動画中の各時刻に対するシーンを識別

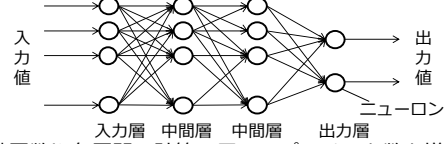


## 時間変化の識別のためのニューラルネットワーク



## 深層学習

- 多階層のニューラルネットワークを用いた機械学習



- 階層数や各層間の計算に用いるパラメータ数を増やすことにより、複雑なモデルを表現可能
- 画像、動画画像等の分類への適用が盛んに研究されている
  - 多量の学習用データを用い、ニューラルネットワークを学習
  - 学習されたニューラルネットワークを用い分類

## 研究目的に対する課題とアプローチ

- 課題：
  - 少ない学習用サンプルを用いた場合であっても、適切なシーン識別を可能とするニューラルネットワークの適用方法の確立
  - 多量の学習データを、抽出したい全シーンに対して準備するのは困難

従来研究：多量の学習データを用いることを前提にニューロン数が多いニューラルネットワークを構成

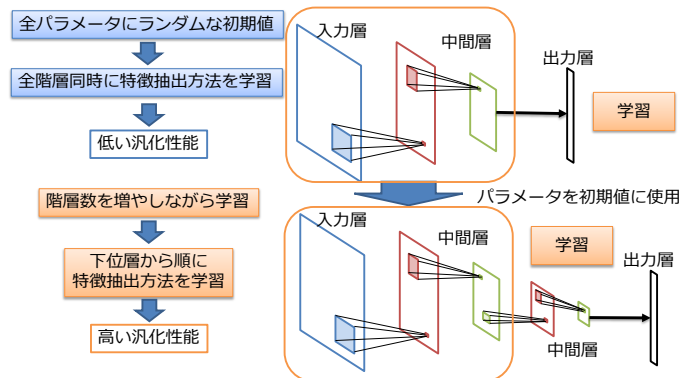
学習対象のパラメータ数が多く、学習に多くのサンプルが必要

- アプローチ
  - シーンの識別にとって重要なフレームのみを入力とするニューラルネットワークを構成

入力層のニューロン数の削減にとまない、中間層のニューロン数も削減

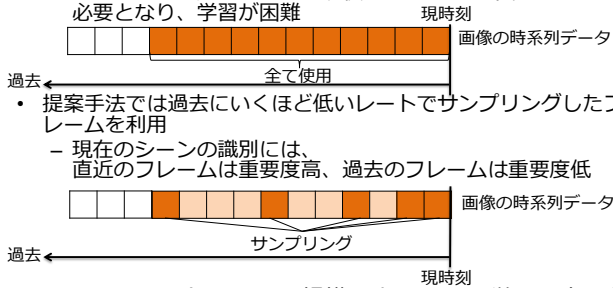
学習に必要なパラメータ数の削減を達成

## ニューラルネットワークの学習方法



## ニューラルネットワークの入力

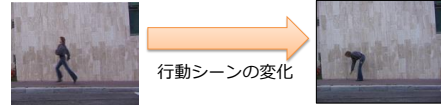
- 過去一定時間内のフレームを利用
  - 全フレームを用いると、大規模なニューラルネットワークが必要となり、学習が困難
- 提案手法では過去にいくほど低いレートでサンプリングしたフレームを利用
  - 現在のシーンの識別には、直近のフレームは重要度高、過去のフレームは重要度低



→ニューラルネットワークの規模を小さくし、学習を容易化

## 評価環境

- 元になるデータセット※
  - 10種の行動パターンが25fpsで撮影された動画
  - 8人が10種の同様の行動をした様子を1動画ごとに収録
- 学習用動画の作り方
  - 8人中6人の、同じ人が行動している動画同士を結合
    - シーン切り替わり後の時刻からサンプリングを開始する
    - シーン切り替わり後の行動を正解ラベルとする
- テスト用動画の作り方
  - 残り2人に対し学習用動画作成時と同様に結合



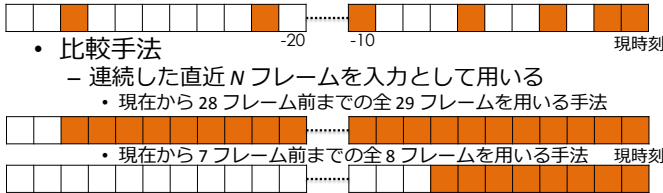
※ <http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>

## 評価方法

識別の正確さとして精度を以下のように定義

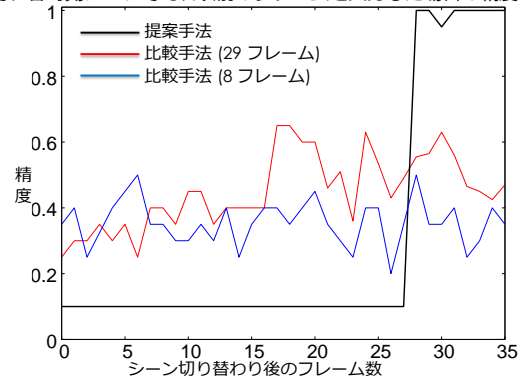
$$\text{精度} = \frac{\text{データに対して正しく識別した回数}}{\text{入力した総サンプル数}}$$

- 提案手法
  - 現在から過去のフレームに向かって間隔を1フレームずつ増やしながらサンプリングして1秒間程度に渡って計8個のフレームを入力として用いる
- 比較手法
  - 連続した直近Nフレームを入力として用いる
    - 現在から28フレーム前までの全29フレームを用いる手法
    - 現在から7フレーム前までの全8フレームを用いる手法



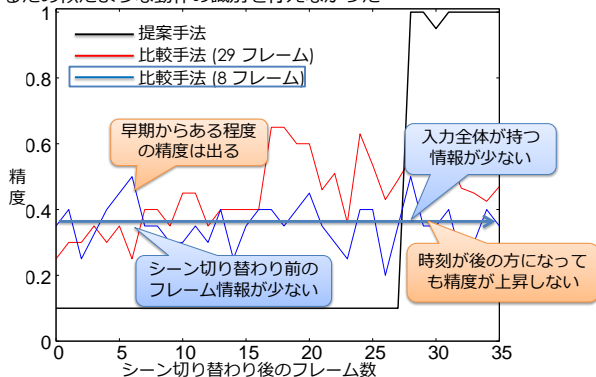
## 評価結果

- 繋ぎ合わせた動画の元の2つの動画シーンが始まるフレームを0として、各時刻についてそれ以前のフレームを入力した場合の精度を示す



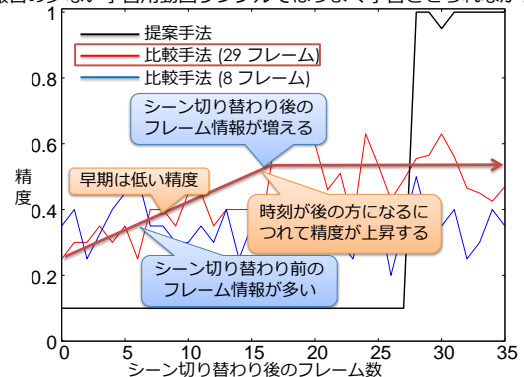
## 評価結果

- 直近8フレームを入力に用いると、入力全体が持つ動きの差分が小さすぎるため似たような動作の識別を行えなかった



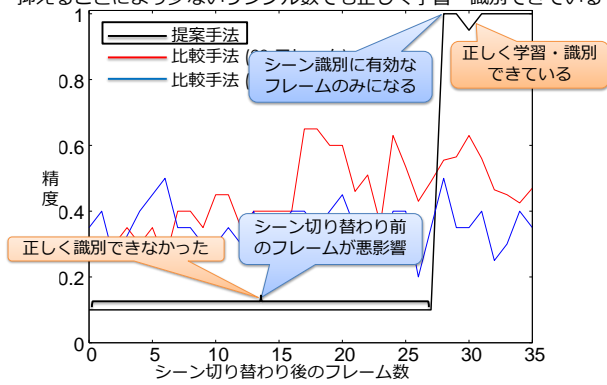
## 評価結果

- 直近29フレームを入力に用いるのは、必要なパラメータ数が多くなり本報告の少ない学習用動画サンプルではうまく学習させられなかった



## 評価結果

- 提案手法は入力全体の持つ動きの差分が大きく、全体のニューロン数を抑えることにより少ないサンプル数でも正しく学習・識別できている



## まとめと今後の課題

- まとめ
  - 深層学習による動画中の各時刻におけるシーン抽出手法を提案した
  - 重要なフレームのみを入力に用いることでニューロン数の少ないニューラルネットワークを構成した
  - 少ないサンプル数での学習・識別を実現した
  - 時系列データを一区間の全データではなく部分的に入力に用いる手法の有効性を明らかにした
- 今後の課題
  - 画像以外の時系列データへの深層学習の応用
    - トラヒックの時系列データをニューラルネットワークの入力に用いたネットワーク制御手法など