

# 映像認識の教師なし学習事例

参考論文：『Look, Listen and Learn』

written by Relja Arandjelovic and Andrew Zisserman (DeepMind)

小野 藍

# 映像認識の教師なし学習事例

## Contents:

1. 従来の映像認識：教師あり学習
2. 最新の映像認識：教師なし学習
3. まとめ

補足資料. 事例のネットワーク構造図

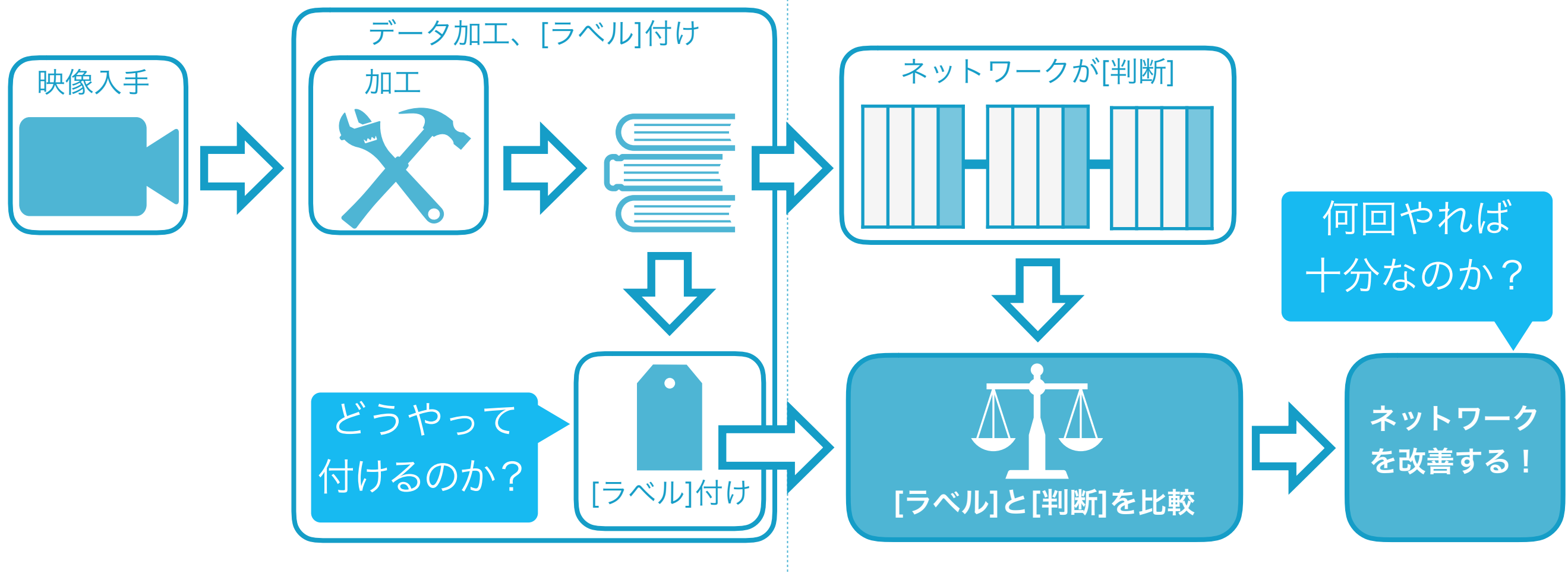
# ネットワーク学習



# 1. 従来の映像認識：教師あり学習(2/2)

学習データ作成

ネットワーク学習



■映像のラベル付けは、自動化・定型化がほぼ不可能

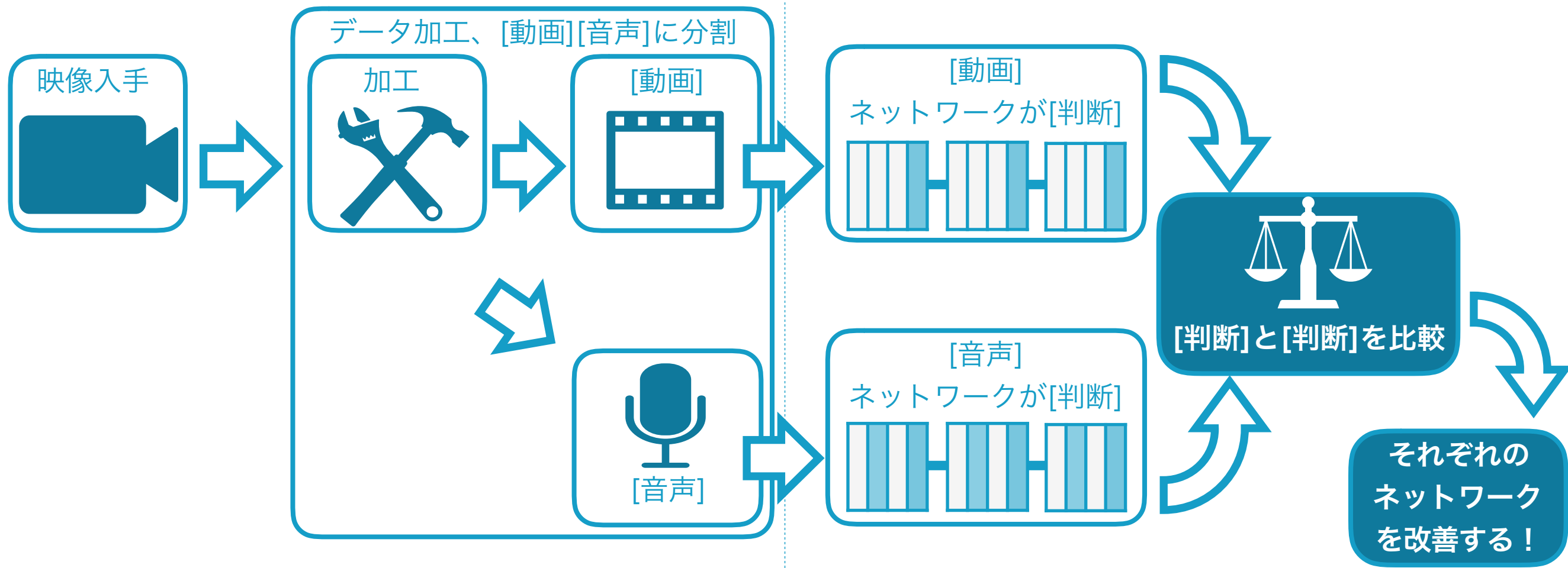
■ネットワーク学習は、数万回、数十万回、あるいはそれ以上の回数が必要

⇒ 映像認識の教師あり学習は、十分な数の学習データ確保が難しい

## 2. 最新の映像認識：教師なし学習(1/2)

学習データ作成

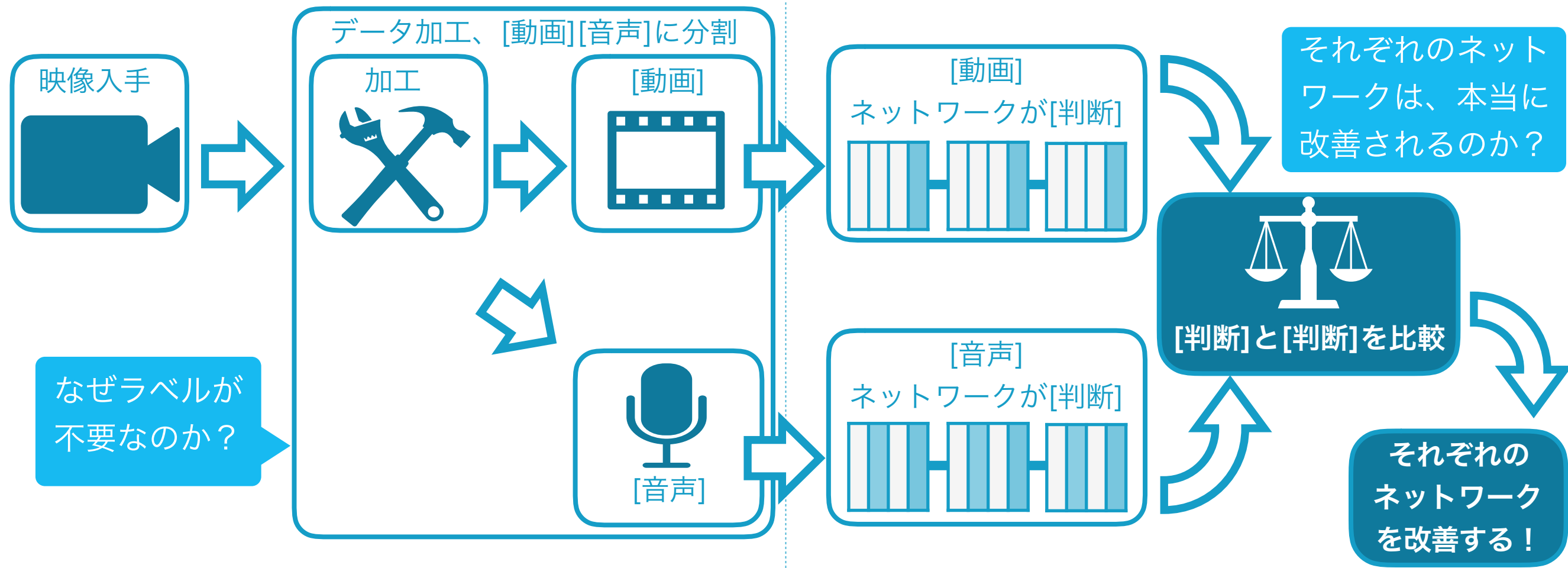
ネットワーク学習



## 2. 最新の映像認識：教師なし学習(2/2)

学習データ作成

ネットワーク学習



- 『[動画]と[音声]の発生源は同じなので、相互に情報を補完しあえる』という考え方
  - 両方のネットワーク共、単体で教師あり学習した他の研究と比べても精度に遜色なし
- ⇒ 教師なし学習で、全ての音声付き映像が学習データになる

### 3. まとめ

従来の映像認識には、十分な数の学習データ確保が難しい、という弱点があった

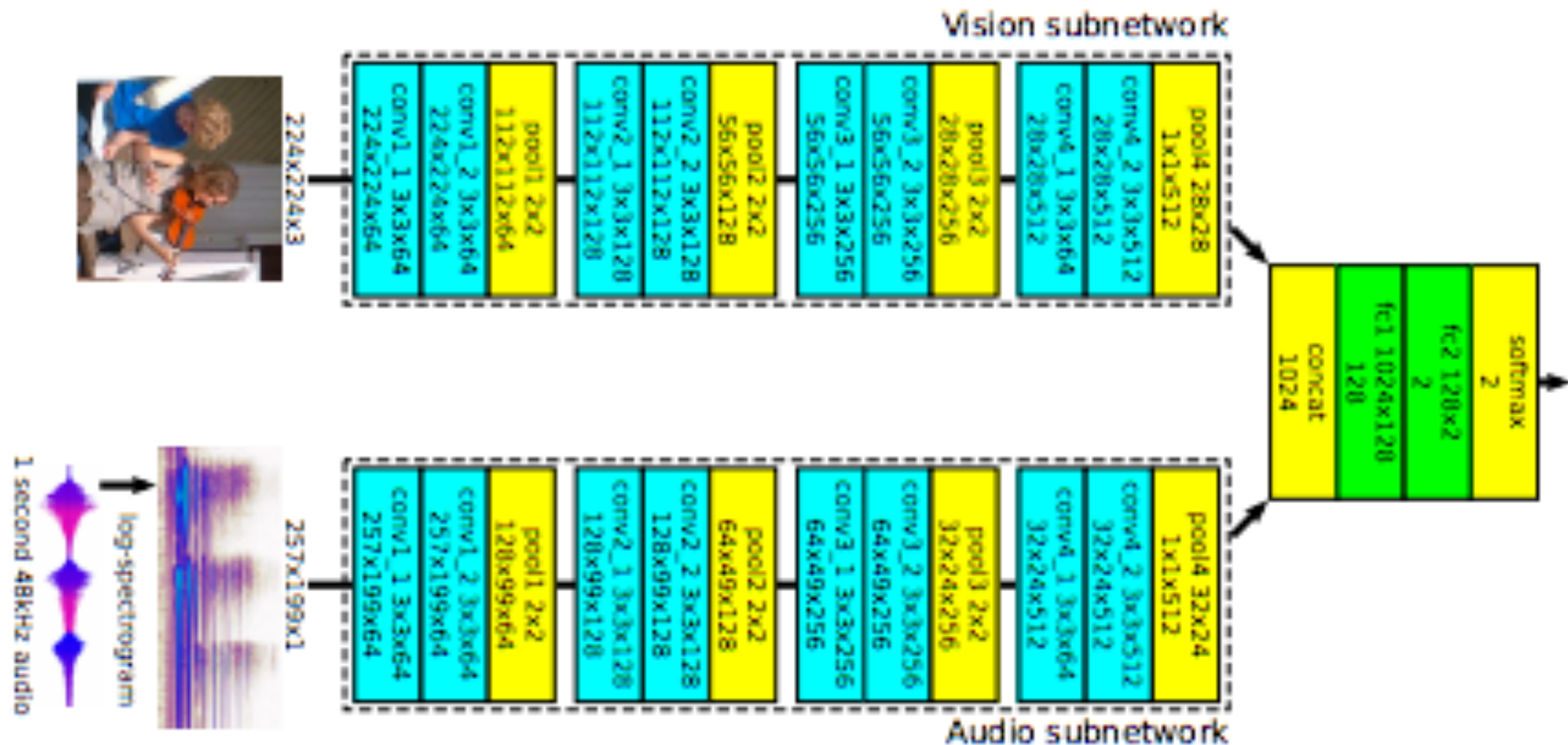


事例の教師なし学習により弱点は克服され、  
全ての音声付き映像が学習データになる



映像認識の精度は、今後加速度的に向上する

# 補足資料. 事例のネットワーク構造図



実装方法はもちろん、なぜこれでネットワークが改善していくのか、現時点では皆目理解できない。



ご清聴有難うございました

小野 藍