

Face Detection and Identification using YOLO (YOLOを用いた顔検出および認識)

著者 MS Antony Vigil , Manasi Makarand Barhanpurkar, NS Rahul
Anand , Yash Soni, Anmol Anand

所属 Department of Computer Science and Engineering SRM
Institute Of Science And Technology Chennai, India

知的高性能計算研究室
M1 邵舒揚 7月8日

概要

- 研究背景
- 提案手法の例を考える
- 全結合層(Fully Connected layer)
- YOLO検出システム
- BOUNDING BOXの回帰
- EYE-SPY
- WordTree
- これからの仕事

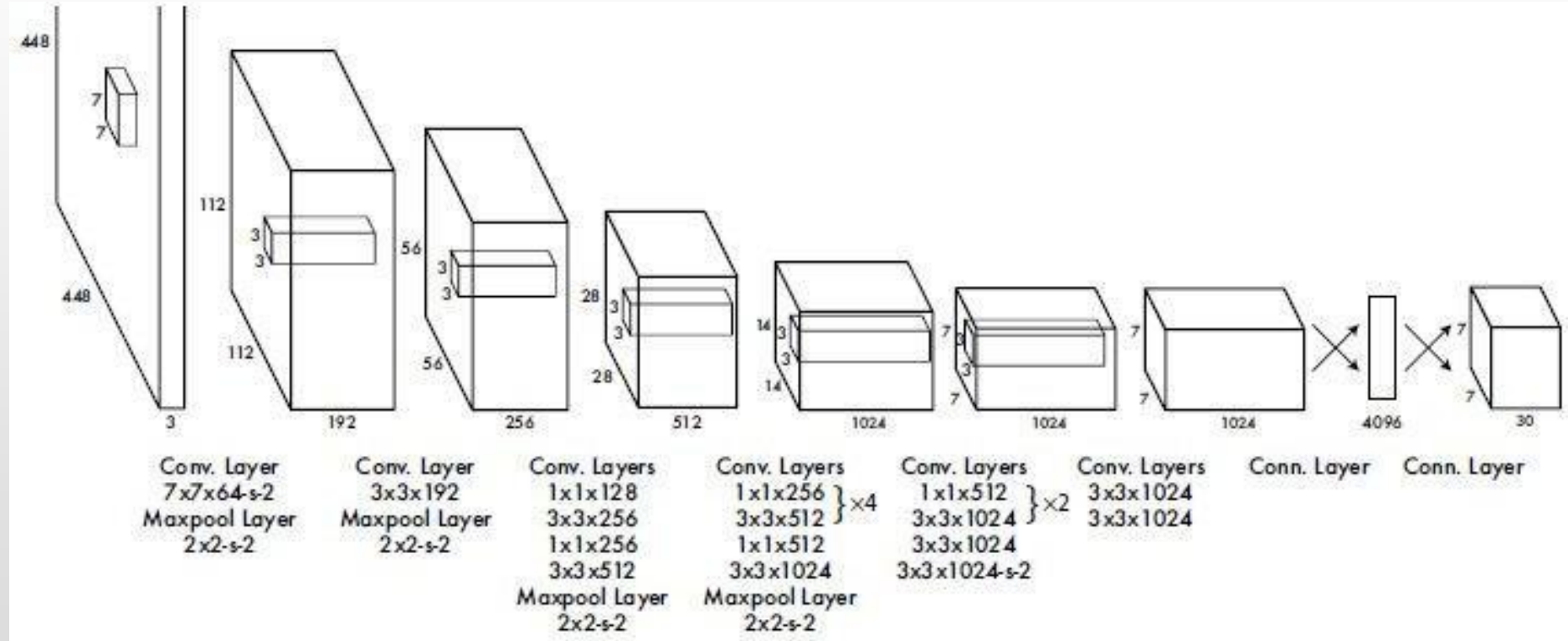
研究背景

- 本文はyoloの応用とその長所と短所について述べた。さまざまなサイズで動作するマルチスケールトレーニング方法を使用して、関係の速度と正確性を向上させます。
- これまでYoloはオブジェクト検出のみを使用してきましたが、これを使用した顔認識は、異なるサイズ、照明、姿勢の人の顔を含む異なるデータセットを使用して精度を向上させることで行われます。



提案手法の例を考える

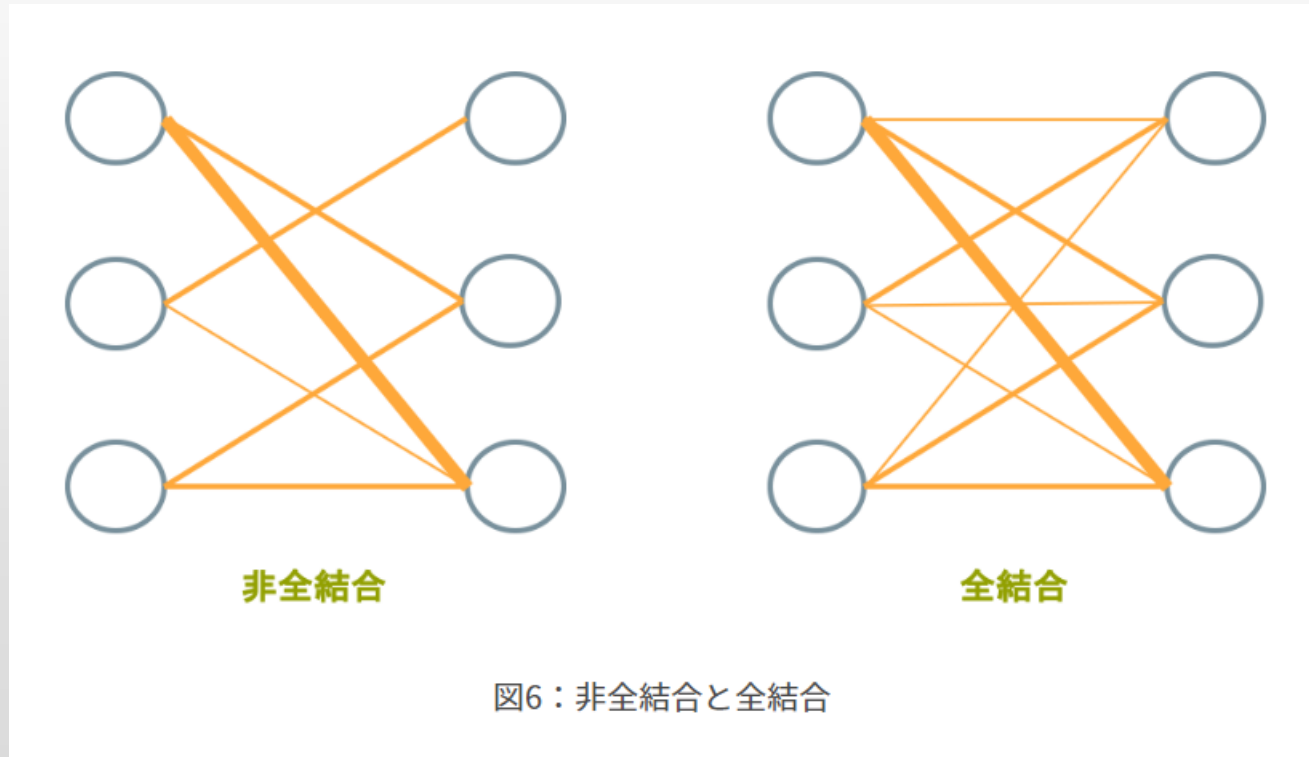
カメラ入力に対する本質的に小さい物体の解像度を特定することはまだ困難です。本研究の目標は、最も遠くにいる人でもアルゴリズムを改善して正確に識別する。



24の畳み込み層と2つの全結合層を使っています。畳み込みに対して半解像度のImage Net分類前訓練を行い、全解像度訓練を行った。

全結合層(Fully Connected layer)

ノードからノードに全て結合するのが全結合で、全てでない手抜きが非全結合です。実は畳み込み層では、全結合を計算すると処理が膨大になってしまうので、非全結合で処理していたのです。



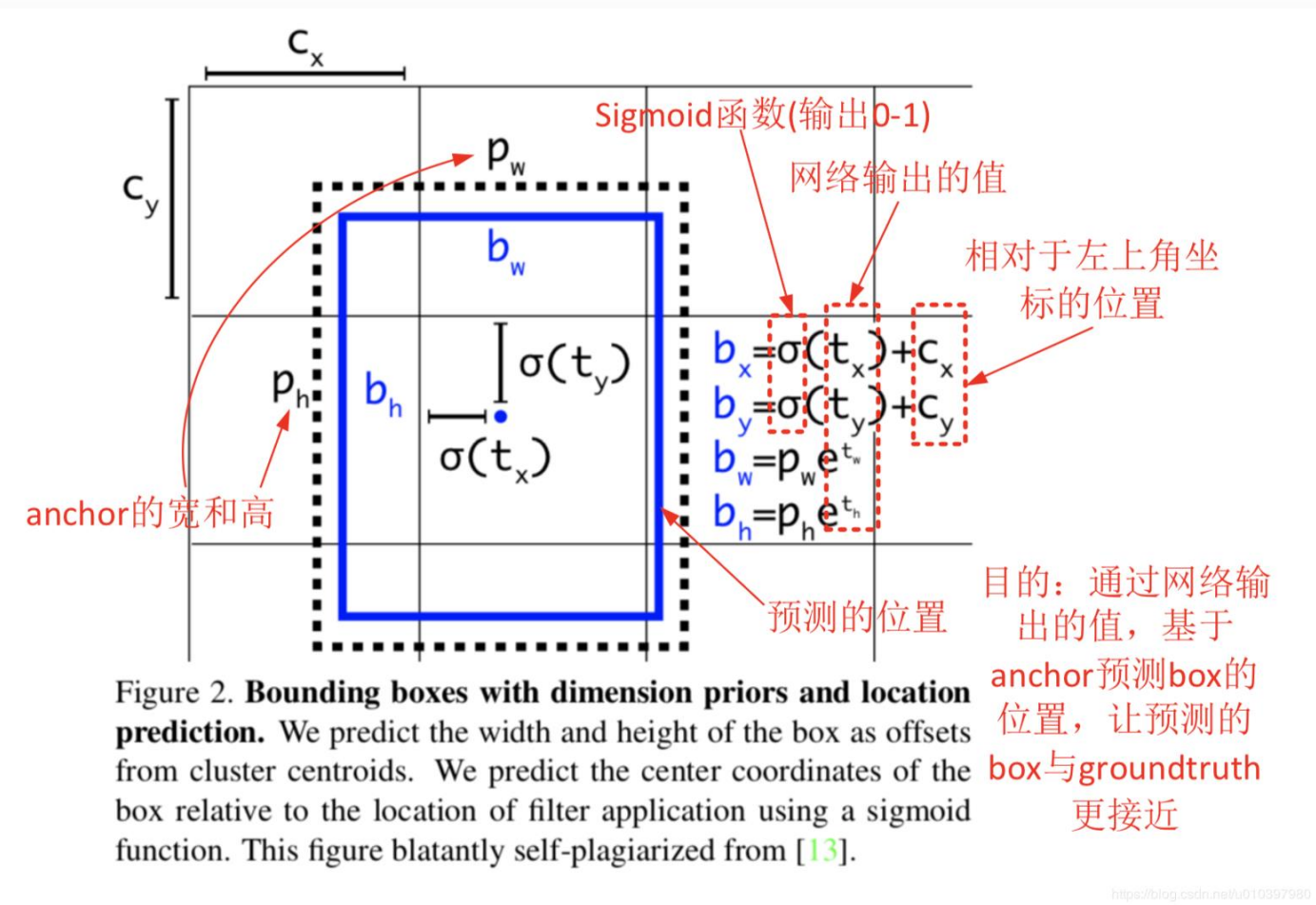
YOLO検出システム



- 1、元の画像を等サイズのグリッドに分割します。
- 2、それぞれのグリッドに関して、事前に定義した形状の、グリッドの中心を中心位置とするバウンディング・ボックスを、事前に設定した数だけ予測します。それぞれの予測には、クラス確率と物体信頼度（その領域が物体を含んでいるか、あるいは背景のみか）が関連付けられます。
- 3、最後に、高い物体信頼度およびクラス確率が関連付けられたバウンディング・ボックスを選択します。最も高いクラス確率を持つ物体クラスが、その物体のカテゴリとなります。

BOUNDING BOXの回帰

下図は、バウンディングボックスの予測式を示しています。バウンディングボックスを予測する場合、 c_x, c_y は左上隅からの相対位置であり、各セルの長さを1とすると、この図では、 $c_x=1, c_y = t_x$ と t_y はそれぞれ0と1の間のオフセットのためのシグモイド出力を通過し、 c_x, c_y と合計されて得られます。 p_w, p_h は手動で設定されたアンカーの幅と高さです。 t_w, t_h をそれぞれ p_w, p_h とすると、bounding boxの幅と高さが得られます。これにより、予測されたbounding boxの x, y, w, h が得られます。予測された x, y, w, h をgroundtruthに近づけることが目標なので、LOSSを書けるようになりました。



EYE-SPY

リアルタイムの顔検出の最も意欲的な応用の1つをご紹介しますが、最高の機械学習を使って、私たちのリソースを使って人間を検出し、行動し、予測します。機械学習は今日の技術進歩のピークであり続けている。最新のテクノロジーと、最も信頼性が高く基本的なカメラ・アプリケーションの1つ、顔検出、目標を組み合わせる。Eye-Spyは最先端、リアルタイム、人間的な検出システムである。すでに流行りの技術YOLOの技術支援を利用して、このアルゴリズムを現代の監視システムに実現しようと計画している。

TECHNIQUES FOR Eye-Spy

顔採取技術を使って、一つ一つの顔を鮮明に覚えておく必要があります。顔採取に使える技術は4つある。

1. 伝統的な
2. 皮膚テクスチャ分析
3. 3D識別
4. 技術的結合

WordTree

modelのtrainigにdetection用のデータとclassification用のデータを混ぜて使うことによって強度を上げる。
detection用のデータは、dogとかboatというラベルが付与されているが、ImageNetでいうとclassification用のデータにはNorfolk terrierとかYorkshire terrierとかBedlington terrierとか、情報の粒度が異なってくる。

HIERARCHICAL CLASSIFICATION

ImageNetのlabelは、WordNetから引っ張ってきており、例えば、Norfolk terrierはdogの中のhunting dogという種類であるとわかるようになっている。

(WordNetは言語のデータベース)

ただ、木構造ではなく直接的なグラフ構造になっている。
例えば、動物 -> 犬 -> 柴犬というのではなく、動物 -> 柴犬と犬 -> 柴犬というような感じ。

そのため、分類毎にsoftmaxが使えるようになる。

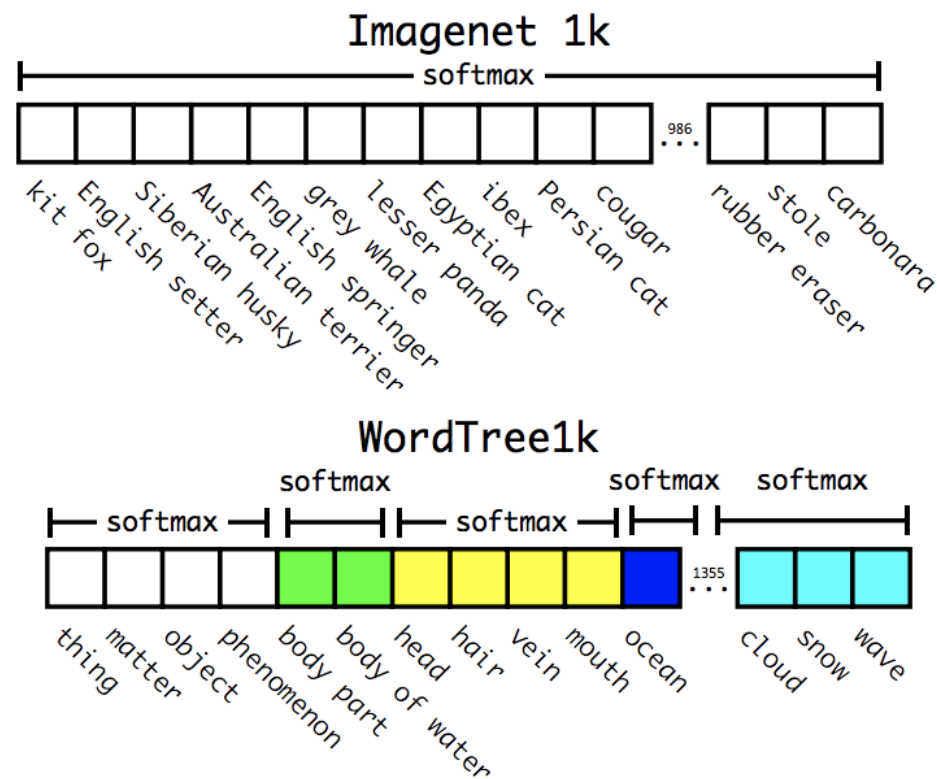


Figure 5: Prediction on ImageNet vs WordTree. Most ImageNet models use one large softmax to predict a probability distribution. Using WordTree we perform multiple softmax operations over co-hyponyms.

このように、WordTreeのノード上で、次のようないくつかの条件付き確率値のノードを計算することができます。テリアこのノードでは、以下の条件付き確率値が得られます。

$Pr(\text{Norfolk terrier}|\text{terrier})$
 $Pr(\text{Yorkshire terrier}|\text{terrier})$
 $Pr(\text{Bedlington terrier}|\text{terrier})$

ノードの確率を予測したい場合は、次のようにWordTreeにしたがって、ルートノードへの条件付き確率を順番に乗算していきます。

$Pr(\text{Norfolk terrier}) = Pr(\text{Norfolk terrier}|\text{terrier})$
 $*Pr(\text{terrier}|\text{hunting dog})$
 $*\dots*$
 $*Pr(\text{mammal}|Pr(\text{animal}))$
 $*Pr(\text{animal}|\text{physical object})$
 $Pr(\text{physical object}) = 1$

WordTreeの原理を紹介した後、WordTreeを使って分類と検出のデータセットを融合させることができます。右図はCOCOとImageNetのデータセットとWordTreeの図です。

WordTreeでは、ラベルノードのCOCOデータセットとImageNetデータセットを区別するために色を使用します。

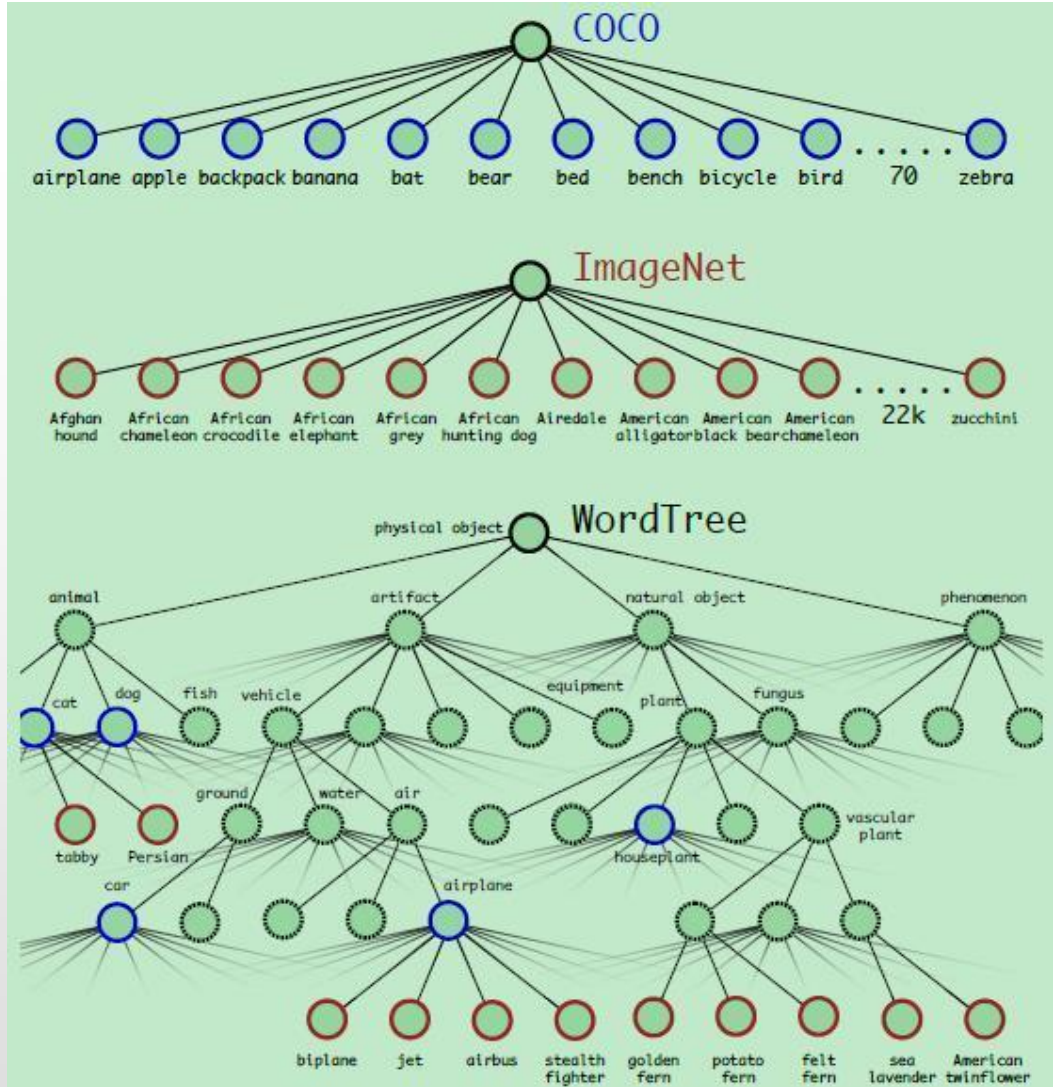


Figure 6: Combining datasets using WordTree hierarchy. Using the WordNet concept graph we build a hierarchical tree of visual concepts. Then we can merge datasets together by mapping the classes in the dataset to synsets in the tree. This is a simplified view of WordTree for illustration purposes. sdl.net/u014380165

これからの仕事

私たちは、視覚世界とその中に存在する顔を、より根拠のある標準化されたモデルにするために、特徴的な情報源や情報構造を結びつけるアプローチを模索し続けていきます。