

変形ARマーカ ～SSD-6D～

ER17076 安井 理

目次

- ① 目的
- ② 概要
- ③ ネットワーク
- ④ トレーニング
- ⑤ 検出

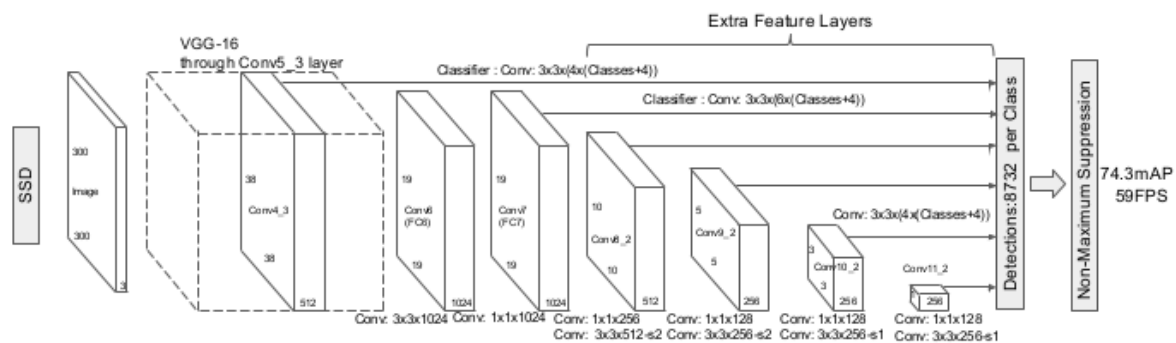
1. 目的

- ・カラー画像の学習のみで高精度かつ高速な学習が行える
- ・SSDの拡張で2Dの検出・6Dの適切な姿勢推定をする
- ・2次元ARマーカを多方向からの読み取りや視点認識につながる

2. 概要

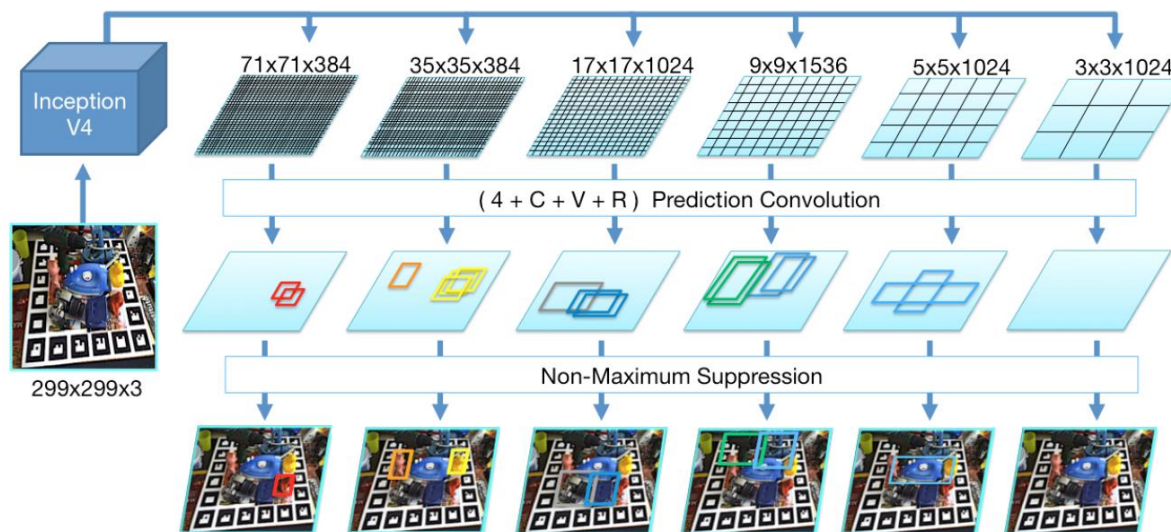
- ・3Dモデルインスタンスの検出と6D姿勢推定
- ・SSDモデルの拡張
- ・トレーニングでは合成3Dモデル情報の使用
- ・モデルの姿勢分解でトレーニングによる対称性の処理
- ・正確な2D検出と6D姿勢生成を行う

2. 概要



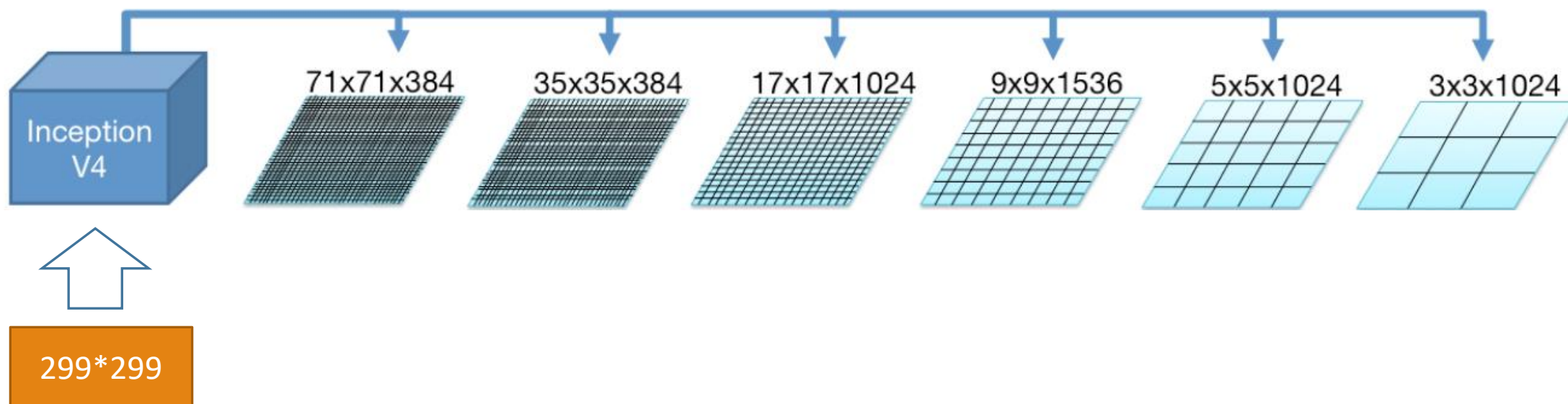
SSD

SSD-6D



3. ネットワーク

- 299×299 のカラー画像を入力とし、事前に訓練されたInceptionV4から
- 複数の縮尺のフィーチャマップを生成
- 前半部は局所的に後半部は大局所的に



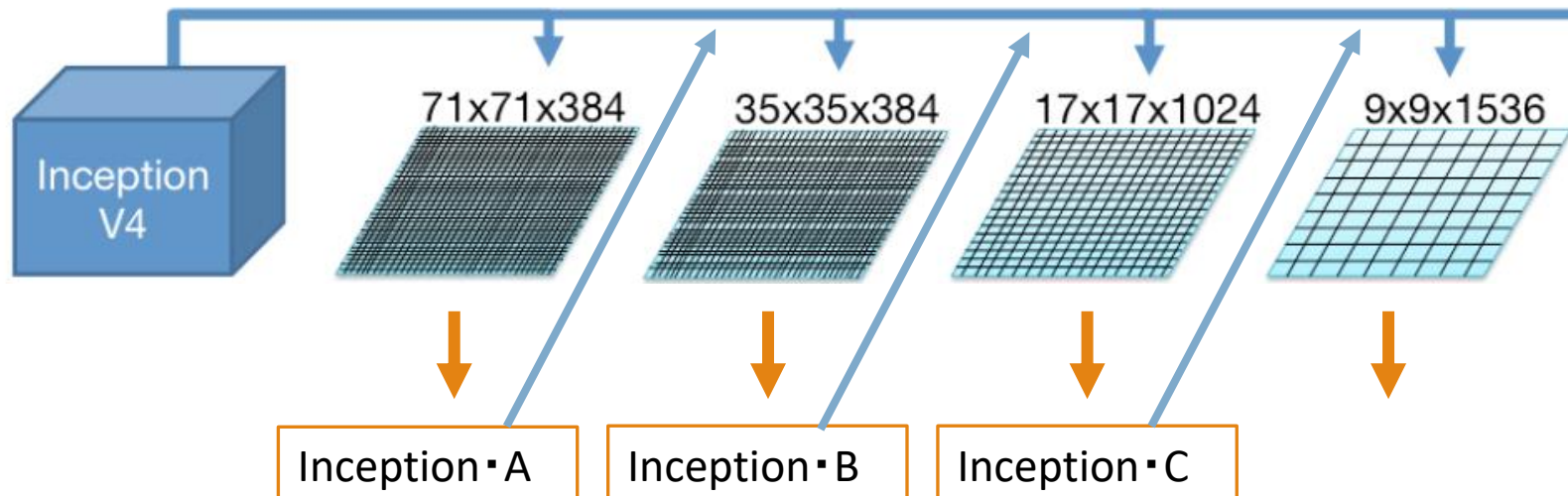
3. ネットワーク

$71 \times 71 \times 384 \Rightarrow \text{Inception} \cdot A$

└ $35 \times 35 \times 384 \Rightarrow \text{Inception} \cdot B$

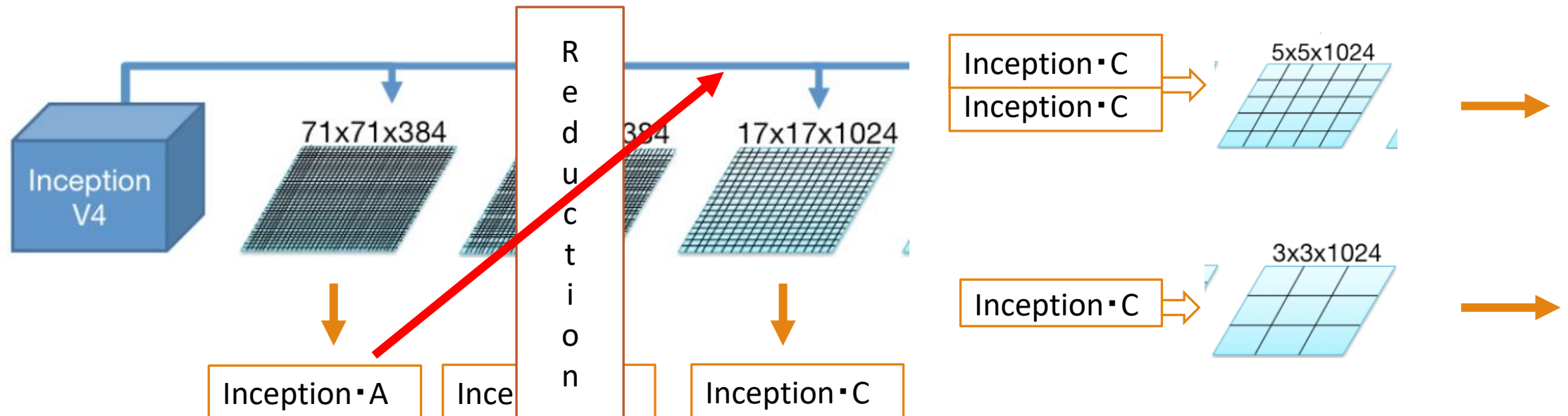
└ $17 \times 17 \times 1024 \Rightarrow \text{Inception} \cdot C$

└ $9 \times 9 \times 1536$



3. ネットワーク

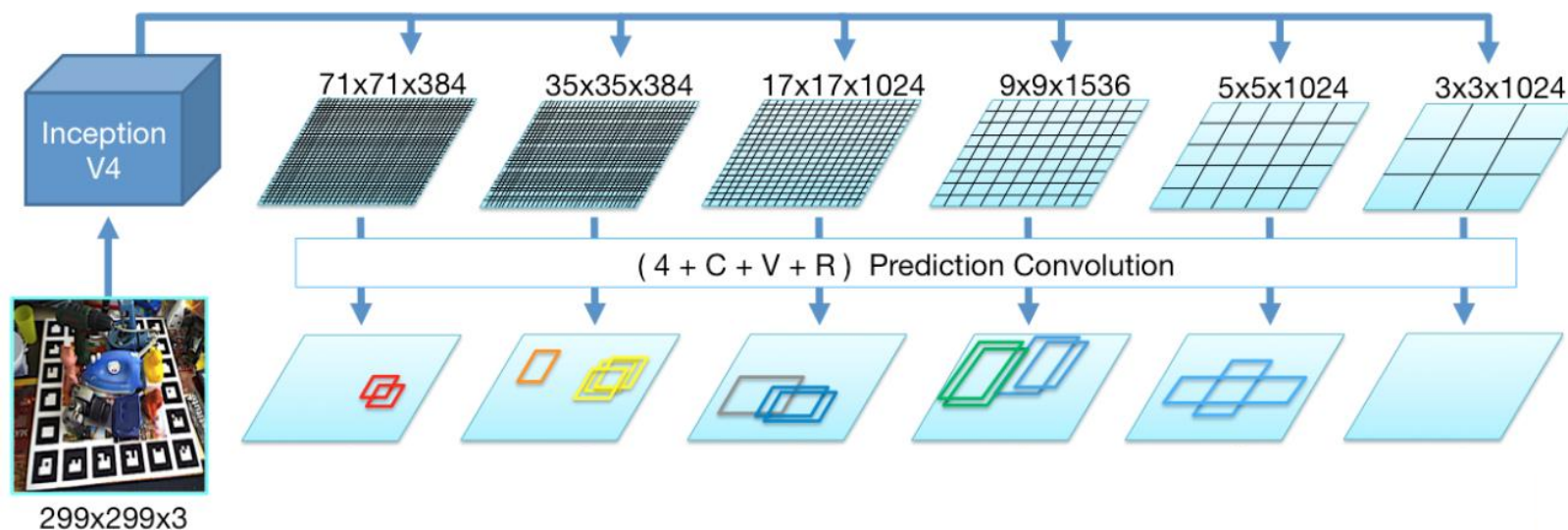
- ・より大きい物体を検出するために2つ拡張する
- ・1つ目:「Reduction・B」の後に「Inception・C」を2度行い5*5*1024のフィルタを通す
- ・2つ目:「Reduction・B」の後に「Inception・C」を行い3*3*1024のフィルタを通す



3. ネットワーク

- ・それぞれの特徴マップをトレーニング済み予測カーネル($C \cdot 4 \cdot V \cdot R$)にかける

C: オブジェクトクラス(分類) 4: バウンディングボックス V: 視点 R: 画内回転



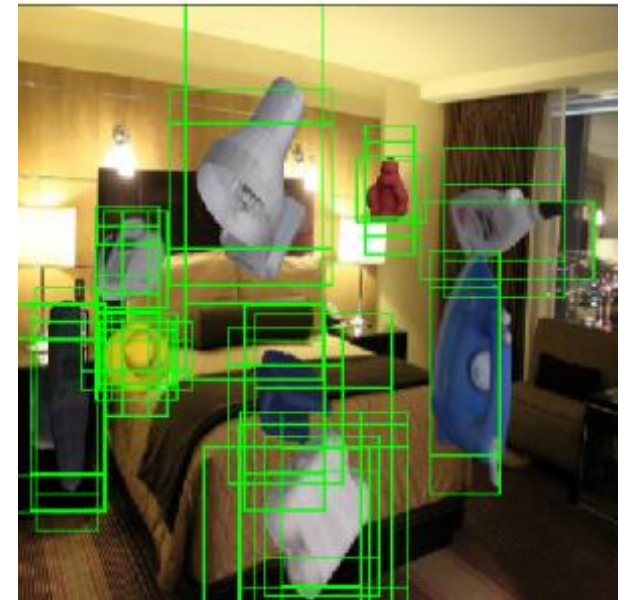
3. ネットワーク

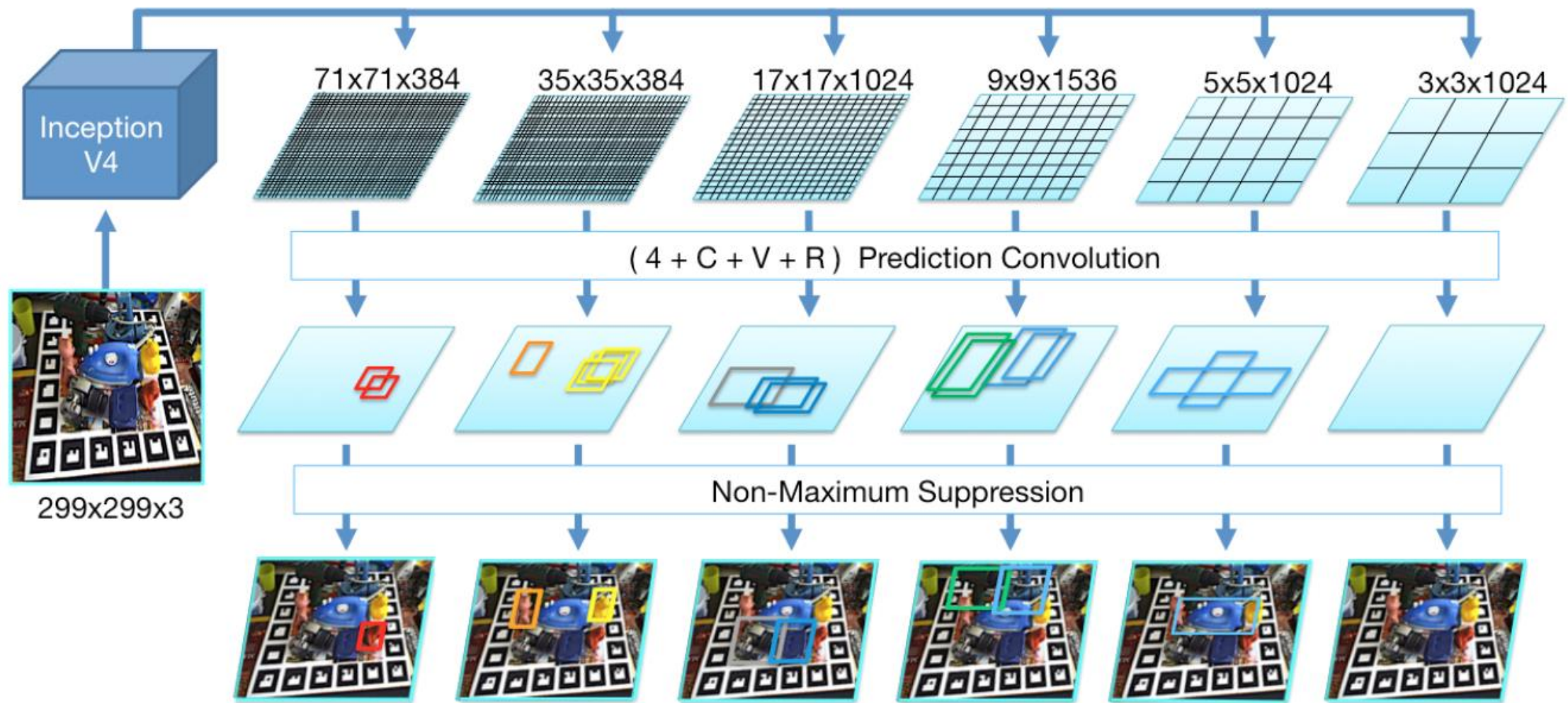
4 (バウンディングボックス)

$C \cdot V \cdot R$ のスコアを求めるための、 $3 \times 3 \times C_s$ (h 高さ w 幅 c チャンネル数)の異なるボックスを用意

$C \cdot V \cdot R$ の一致度合い(IoU?)を計算

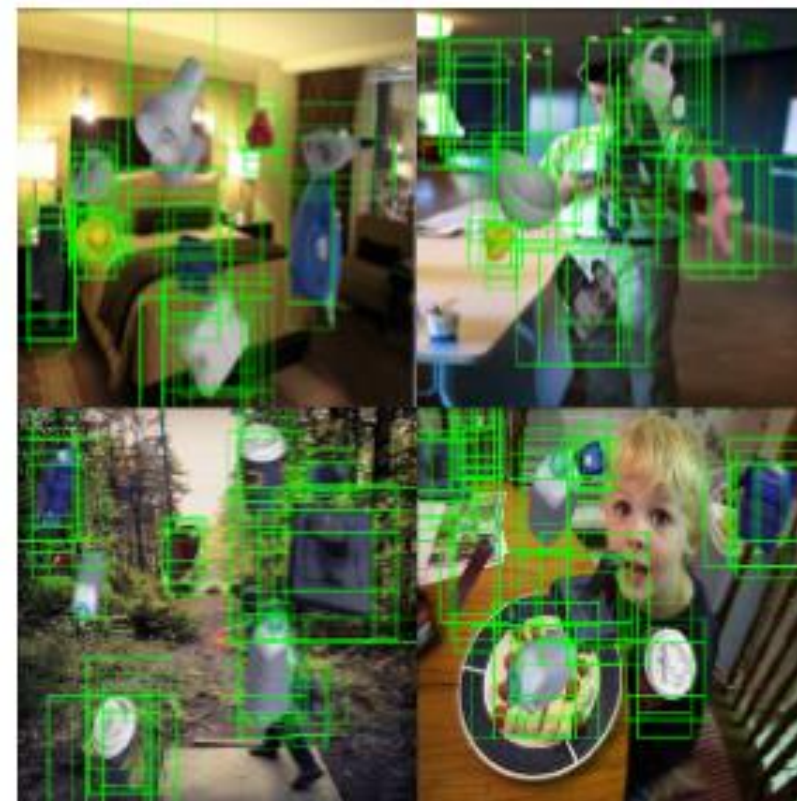
各スケールの特徴マップに対して $a(ws \cdot hs \cdot Bs \cdot (C+V+R+4))$ のboxを生成





4. トレーニング

- ・ 予測カーネルの学習
- ・ MS COCO [20]のランダムな画像を背景として使用
- ・ オブジェクトを画像に張り付ける



4. トレーニング

- ・提示されたボックスのオブジェクトクラス $\text{IoU} > 0.5$ を使用
- ・オブジェクトに対して一致度の最も高いV視点とR画内回転をボックスとして回帰
- ・同じ画像で明るさコントラストなどを変更してトレーニング
- ・逆伝播のプロセスでは、ポジティブ1:ネガティブ2の比率になるようハードネガティブを選択する

4. トレーニング

視点と回転の値を含めるようにSSDの式を拡張

クラス確率(L_{class})の正と負を合計

正のboxは、ビューポイント(view)と画内分類(inplane)の重み付け
ボックスコーナのフィットエラー

この3つは、標準のソフトマックスエントロピー損失を使用

$$L(Pos, Neg) := \sum_{b \in Neg} L_{class} + \sum_{b \in Pos} (L_{class} + \alpha L_{fit} + \beta L_{view} + \gamma L_{inplane}) \quad (1)$$

5. 検出

- ・特定のしきい値を超えるすべての検出し収集その後、非最大抑制(エッジ検出)を実行
- ・すべての視点・画内回転・オブジェクトIDとスコアの付いた2Dbboxが生成
- ・視点と画内回転を解析することで6D推定が可能になる