

画像認識 1 章

tattaka

画像認識 (image recognition) とは

画像に写る内容を理解すること



例えば、上の写真には車と家が写っている

-> 画像を入力として [“車”, “家”] などを出力する

画像認識は難しい



図だと [“車”, “家”] でも [“フォルクスワーゲン”, “木”, “家”] でも
どちらでも正しい

- ▶ 注目する場所, 認識に必要な前提知識, クラス自体の曖昧性
によって結果が変わる

画像認識の歴史

- ▶ もともと**知能ロボットの視覚を実現するため**に研究がスタート
- ▶ 1970~80年代では人間のもつ知識をルールベースでプログラム->**信号を記号で表現する**アプローチ
- ▶ 1980年代以降は形や色，テクスチャ，動きなどの幾何学的・物理的な側面に焦点を当て大きな進歩
- ▶ 1999年に局所特徴**SIFT**が開発されてから多様なクラスを認識することが可能に
- ▶ 2009年には大規模データセット**ImageNet**が作成され，2012年に深層学習モデルを用いたアルゴリズムが既存手法を大きく上回った

画像認識の分類

- ▶ **物体認識** (object recognition)
入力画像に写る物体を理解し、適切なラベルを付与する
 - ▶ **インスタンス認識** (instance recognition) 物体の固有な名前を出力
 - ▶ **クラス認識** (class recognition) 物体の属する概念を出力
- ▶ **シーン認識** (scene recognition)
画像が表現する包括的な状態を理解する

これらに関連した分野として

- ▶ **画像アノテーション** (image annotation)
狭義では複数のラベルが付与されたデータ構造を用いて入力画像に複数のラベルを付与する
- ▶ **物体検出** (object detection)
インスタンスやクラスだけではなく、物体領域を推定する
 - ▶ **セマンティックセグメンテーション** (semantic segmentation)
物体とその他の境界を切り分けて推定する などがある

クラス分類の手順 1

1. サンプルング (sampling)

入力画像の代表的な点を抽出すること

▶ 疎なサンプルング (sparse sampling)

画像内のエッジやコーナなどの特徴的な点を抽出する
特徴するためのアルゴリズムを**検出器** (detector) と呼ぶ

▶ 密なサンプルング (dense sampling)

一定の画素ごとに代表点を抽出する の 2 つに分けられる

2. 局所記述

獲得頂点の周りに小さな領域 (画像パッチやパッチと呼ぶ) を設定し、何らかの方法でパッチから特量を抽出する
得られたパッチの特徴量と検出器から得られる情報合わせて**局所特徴** (local feature) と呼ぶ

3. 統計的特徴抽出

局所特徴は外乱などの影響でノイズを含むので特徴群の確率統計的構造に用いて処理すること
主に主成分分析やフィッシャー線形判別分析などが利用される

クラス分類の手順 2

4. コーディング (coding)

局所特徴を認識に有効なある次元数の特量ベクトルに変換する処理
代表的な方法としてベクトル量子化など
変換したい次元数と同じ数の代表点に局所特徴をクラスタリングする手法

5. プーリング (pooling)

ある画像領域から得られた特徴ベクトル群を 1 本のベクトルにまとめる処理
画像一枚を代表するベクトルを **画像特徴ベクトル** (image feature vector) と呼ぶ
代表的な手法として平均値プーリングなど

6. 分類

画像特徴ベクトルをクラスに割り振る処理これを行うアルゴリズムを **分類機** (classifier) と呼ぶ

シーン認識, 詳細物体認識や画像アノテーションでも上記の手順を利用することが多い

画像アノテーションの場合, 通常のクラス分類とは異なりノンパラメトリックな方法が積極的に活用される

深い構造と浅い構造

ステップ 1~5 を行うことは $H \times W \times D$ の 3 階テンソルを入力して $H' \times W' \times D'$ の 3 階テンソルを出力するシステムといえる

この一連の手順を多段に重ね繰り返し行い最後にステップ 6 を用いて分類を行う構造を**深い構造**と呼ぶ

またこれに対して、ステップ 1~6 を一回しか行わない構造を**浅い構造**と呼ぶ

物体検出の手順

1. 物体領域候補の抽出

まず画像から物体領域の候補を抽出する

確実な手法として**スライディングウィンドウ法** (sliding window method) が挙げられる

この手法はある決まった大きさの小領域を一定の画素感覚ですらしながら領域候補を抽出する

この手法は膨大な計算コストがかかるため、入力画像中で**物体らしさ**を評価しておいて候補数を絞り込むという方法が有効

2. 物体領域候補の物体認識

1で抽出した領域をクラス認識にかけ、その領域の物体らしさを計算する

認識対象のクラスを**正例**、それ以外を**負例**と呼ぶ

物体検出器は正例とフレイを集めてきて機械学習の手法により訓練される

物体検出では**分類困難な負例**のみを選別して学習に利用する

3. 検出領域の絞り込み

同じ物体の中でも一番もっともらしい領域を絞り込む

インスタンス認識の手順 1

撮影条件に依存せずに**同じものを同じものとして認識**する
手順として大きく画像の登録とクエリ画像の認識の2つに分けられる

1. 画像の登録

1.1 特徴点検出と記述

入力画像から局所特徴群を抽出する
インスタンス認識では**幾何学的不変性**をもつ局所特徴がよく用いられる

1.2 画像特徴の計算とインデックス化

局所特徴を利用して画像特徴を計算し、インデックス化する
インデックス化には画像特徴に量子化手法やバイナリ化手法を用い、特徴が近いものがインデックスも近いようになる手法を利用する

1.3 ルックアップテーブルへの登録

インデックスと画像を対応づけたルックアップテーブルを構築し、データベースに登録する

インスタンス認識の手順2

2. クエリ画像のインスタンス認識

2.1 クエリ画像のインデックス化

1.1 と 1.2 と同じ手順で行う

2.2 クエリ画像のインデックスに対応する候補画像群の抽出

ルックアップテーブルを用いて、クエリ画像のインデックスに対応するデータベース内の画像を抽出する
得られる少数の候補画像群を**ショートリスト**と呼ぶ

2.3 局所特徴の幾何学的一貫性の検証

ショートリストの中には同じインスタンスではないがクエリ画像ときたような画像がある可能性がある
クエリ画像とショートリストとの**局所特徴マッチング**を行い、クエリ画像の局所特徴の空間配置と似ている画像を選別し最終的な検索結果とする

画像認識の認識性能を向上させるための原則

データ処理定理

世界の状態を w 、収集したデータを d 、 d を処理したデータを r とすると d が伝達する w に関する平均情報量 $I(w; d)$ は r が伝達する w に関する平均情報量 $I(w; r)$ よりも大きくなる

これは「データ処理はデータのもつ情報を破壊させるのみで決して増やすことはない」ということである

つまり、認識モデルの性能限界は訓練データにのみ決まってしまう

また、前段のモジュールで有益な情報を得られていなければ後段の処理で性能を上げるのは限界があるため入力に近いモジュールを改善することが重要