

画像が提示されていく時、オブジェクトの繰り返し数をカウントアップしていくと、同じような画像では同じオブジェクトのカウントが多くなり、異なる画像群の提示ではカウントがバラけるでしょう。共通のところが多画像群では、高いカウントのオブジェクト群が沢山あらわれるでしょう。そうして、同じクラスに分類されていくはずで

オブジェクトは基本的に画像の部品の属性群になるでしょう。画像の部品とは、顔画像では、眉毛とか目とか鼻とか口とかですね。そして、部品の属性とは、大きさ、向き、配置が主なものになり、オブジェクトとは、この属性の部品対での相対値によって定義されます。つまり、顔画像では、目と眉の相対的な大きさ、相対的な向き（斜めかどうか）、相対的な配置（真上、真下、右、左とか、距離）ですね。そのように部品すべての組み合わせについて、オブジェクトの分布を観ていくことになります。全ての部品の組み合わせですから、部品数が多いと対応すべきオブジェクトが多くなりすぎますが、部品が多いと、そのクラスタ分割して、オブジェクトが扱える範囲なるように対象部品数を減らします。

大きさは、1次元のマップ、向きは2次元のマップ、配置は2次元のマップでオブジェクトは表現可能です。オブジェクトのマップを一行に並べて、全てのオブジェクトの相対値をプロットしていくと、2つの画像の相同性が各オブジェクト毎に把握できるようになります。その相同性の高い部分を評価することによって、画像の分類ができるでしょう。

オブジェクトマップのシーケンスを各画像が提示される毎に記憶保存しておき、あとで一括してマップでの近さを分類していく方法と、オブジェクトマップの値の近いと判断する範囲をあらかじめ定めて置き（量子化）、画像が提示される毎に近い値のオブジェクトマップのカウントをアップしていく方法があります。カウントが高いマップが一群の画像の特徴となります。

この相同性の高いものを一つのクラスの画像と見なすことにより、文字認識やシンボル認識ができるようになります。カウントの高いマップがオンの状態、そうでないマップがオフの状態という入力ノードと画像のクラスを出力ノードとするニューラルネットワークを作れば、クラスを学習するディープラーニングシステムになります。