Fine-Grained Classification

What is fine-grained Classification?

- Classification with categories that are very similar.
  - Can think of these as subcategories.
- Examples:
  - Face recognition
  - Species ID: determine plant species from leaf, breed of dog, species of bird.
  - ID menu item at McDonalds
  - Determine type of shoe.
- Categories are different, but share a common part structure.

Rosch – basic level categories

- Rosch hypothesized that there's a *basic* level category that is recognized first.
  - Furniture, *chair*, rocking chair.
  - Animal, *dog*, beagle
  - Tree, *oak*, white oak
- Purpose of categories is enable inferences (tigers are dangerous).
  - Informative category has many co-occurring attributes
  - But if category is too fine-grained, we can't generalize and they're hard to identify, since they share attributes with other categories.

Evidence for Rosch

- Children learn to name these earlier.
- Highest level category that is visually distinctive.
- In naming attribute, few for superordinate categories, many for basic level, few additional for subordinate level.
- …

Fine-Grained Classification

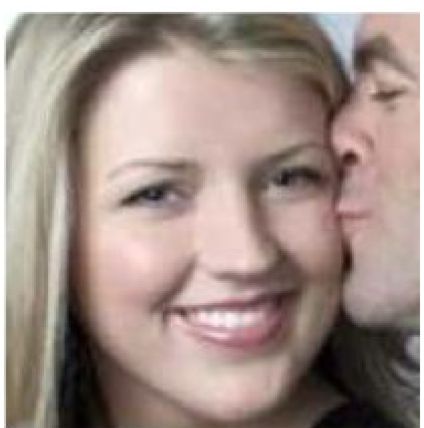
- Classification can be defined as identification at the basic level.
- Fine-grained classification is at a subordinate level.
  - Instances from the same basic level category
  - Share parts == visually similar.

Face Recognition

- Tasks
  - Recognition: Training set, gallery, probes
  - Verification
- Holistic methods in 90s.
  - Eigenfaces
  - Fisherfaces
  - Same vs. different
    - Joint Bayesian
    - Using classifier
- Descriptors
  - Direction of gradient
  - Local binary patterns
  - SIFT-based features

Alignment

- Early methods implicitly assumed images well aligned
  - All frontal
- Unconstrained data sets required explicit methods for alignment.



Alignment

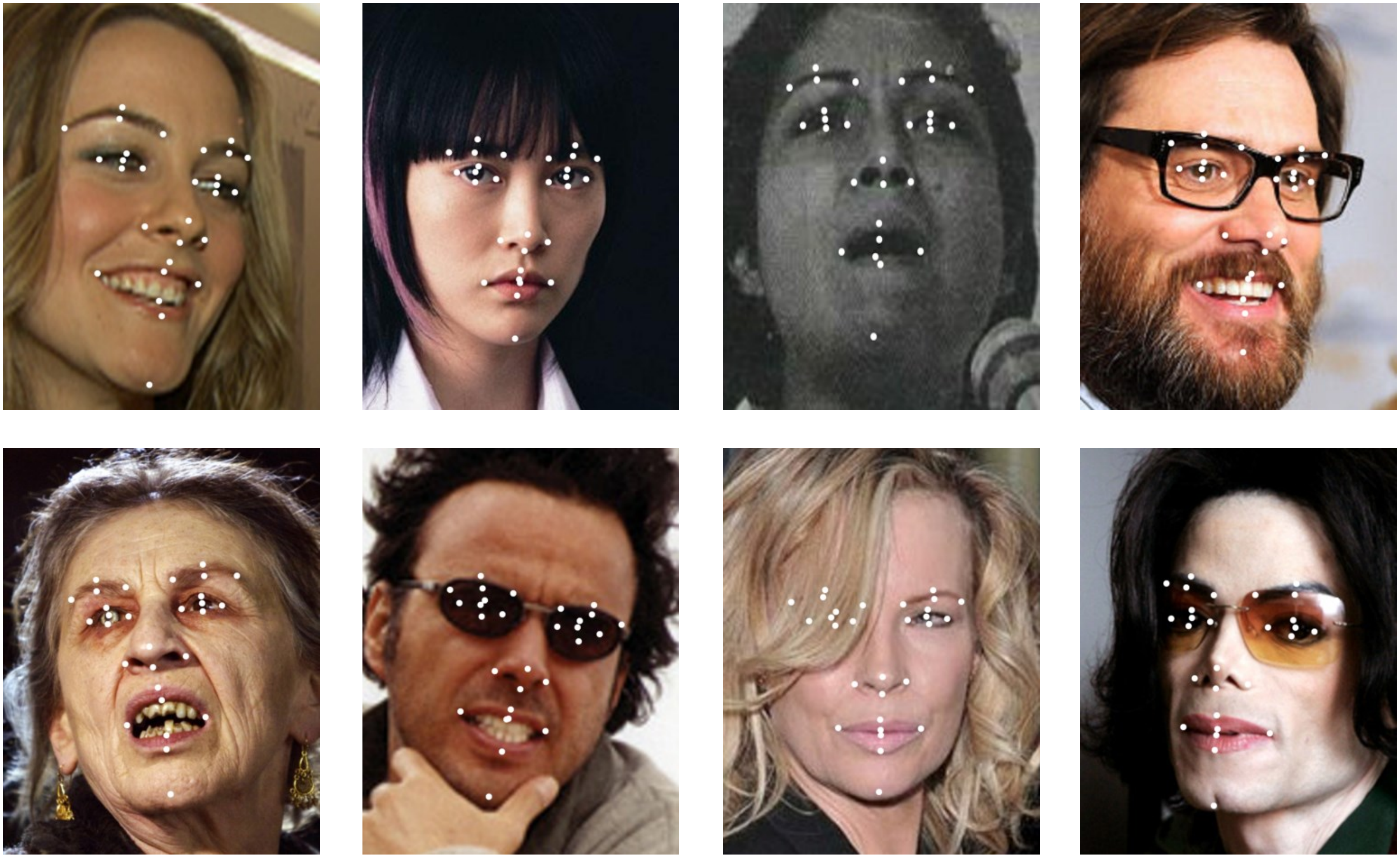
- Find fiducial points
- Transform image to canonical position
  - Similarity transformation
  - 2D warp
  - 3D rotation

Consensus of Exemplars for Part Detection

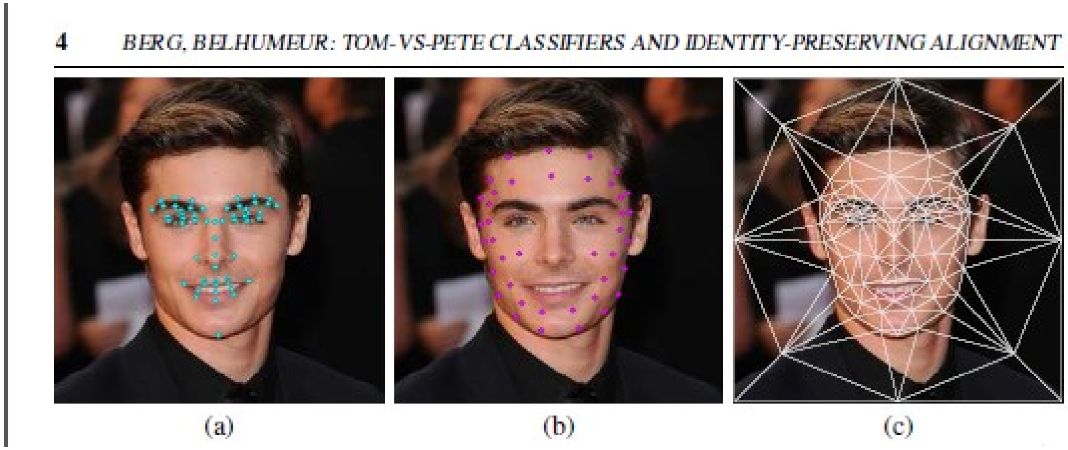
- Bayesian formulation combines prior and image data
- Bottom-up local part model as image data term
  - Sliding Window SVM
- Non-Parametric prior model of part configurations
  - Prior consists of thousands of labeled examples
  - In principle, integrate over all possible configurations, finding MAP estimate combining prior and data.
  - In practice, use RANSAC and sum over most likely configurations and poses.

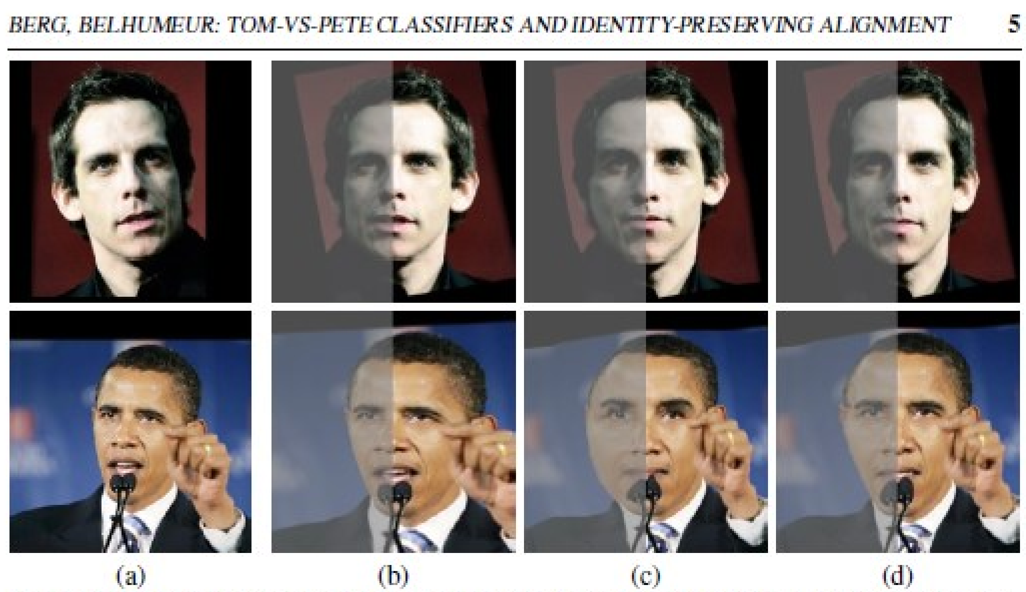
(Belhumeur, Jacobs, Kriegman, Kumar, 2012)

Qualitative Results



Identity-Preserving Warps





DeepFace

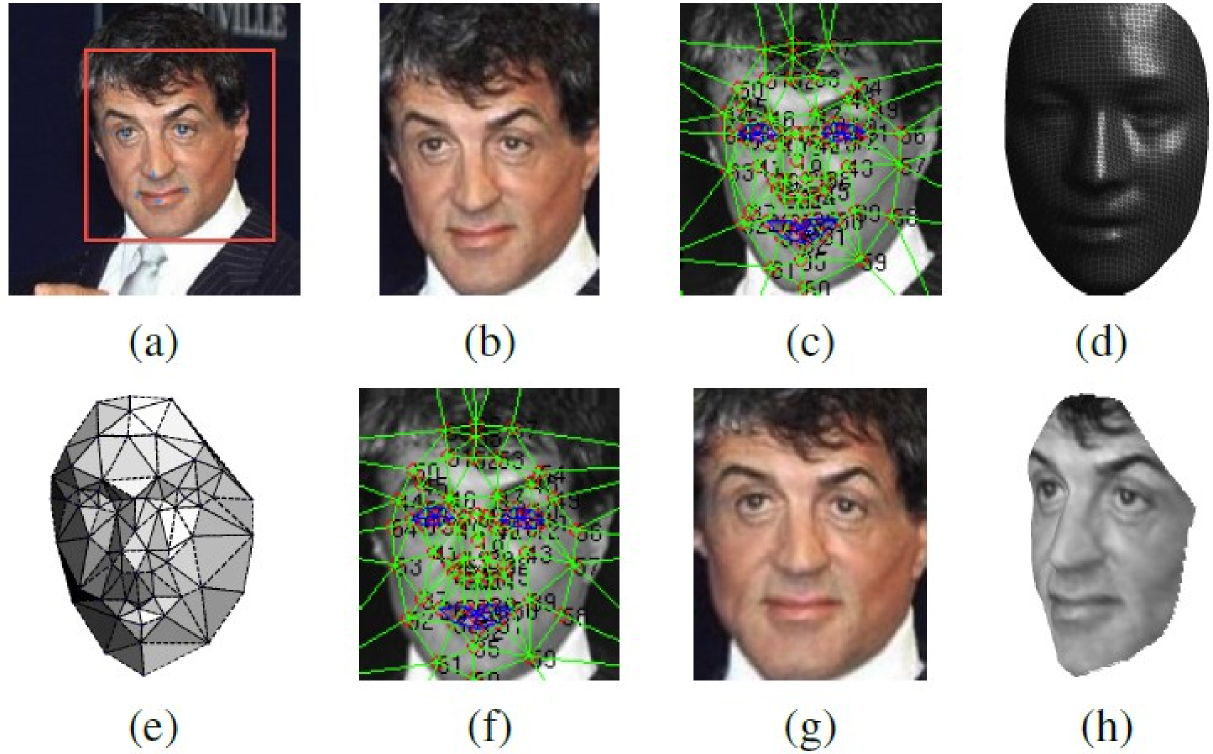


Figure 1. **Alignment pipeline.** (a) The detected face, with 6 initial fidu-

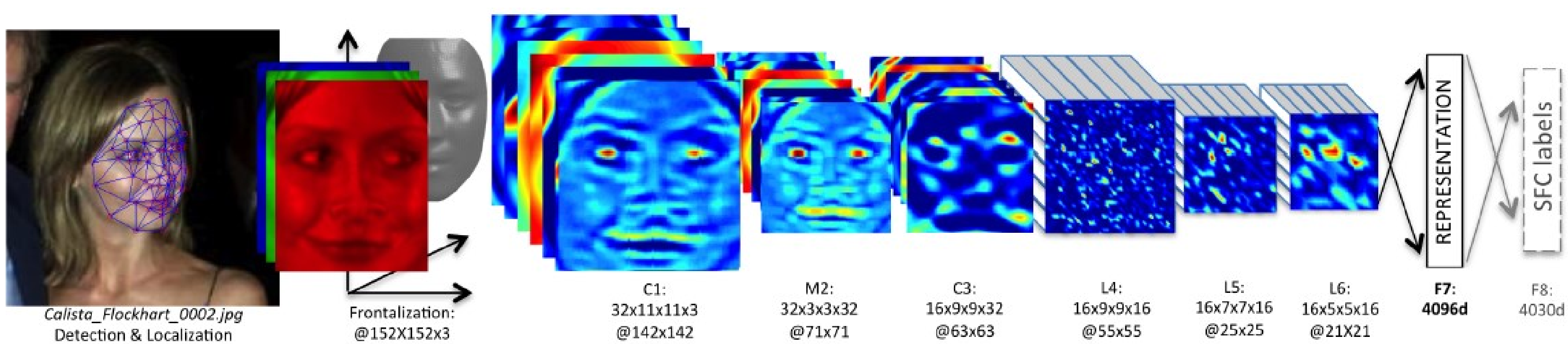


Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

Food 101

Fig. 2: Here we show one example for 100 out of the 101 classes in our dataset. Note the high variance in food type, color, exposure and level of detail, but also visually and semantically similar food types.

Cars

Figure 3: One image each of 196 of the 197 classes in car-197 and each of the 10 classes in BMW-10.

Dog Breed Classification Using Part Localization

Jiongxin Liu[1], Angjoo Kanazawa[2], David Jacobs[2], and Peter Belhumeur[1]

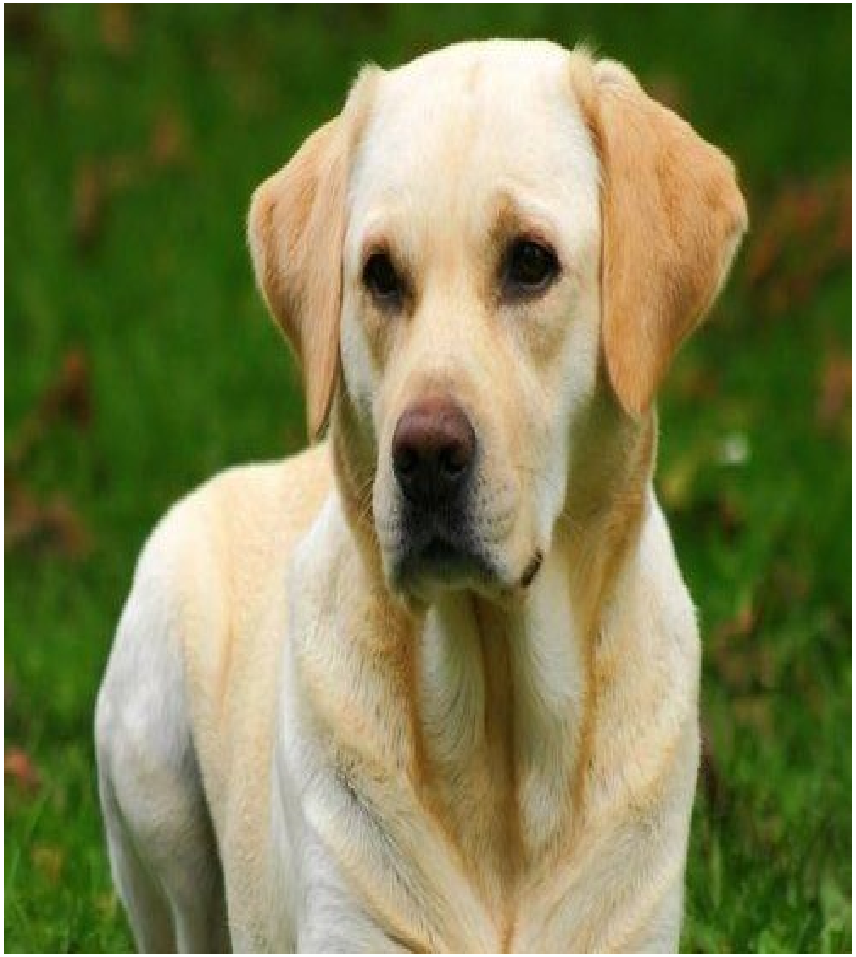
[1] Columbia University [2] University of Maryland

Low inter-breed variation

Norfolk Terrier or Cairn Terrier?

High intra-breed variation

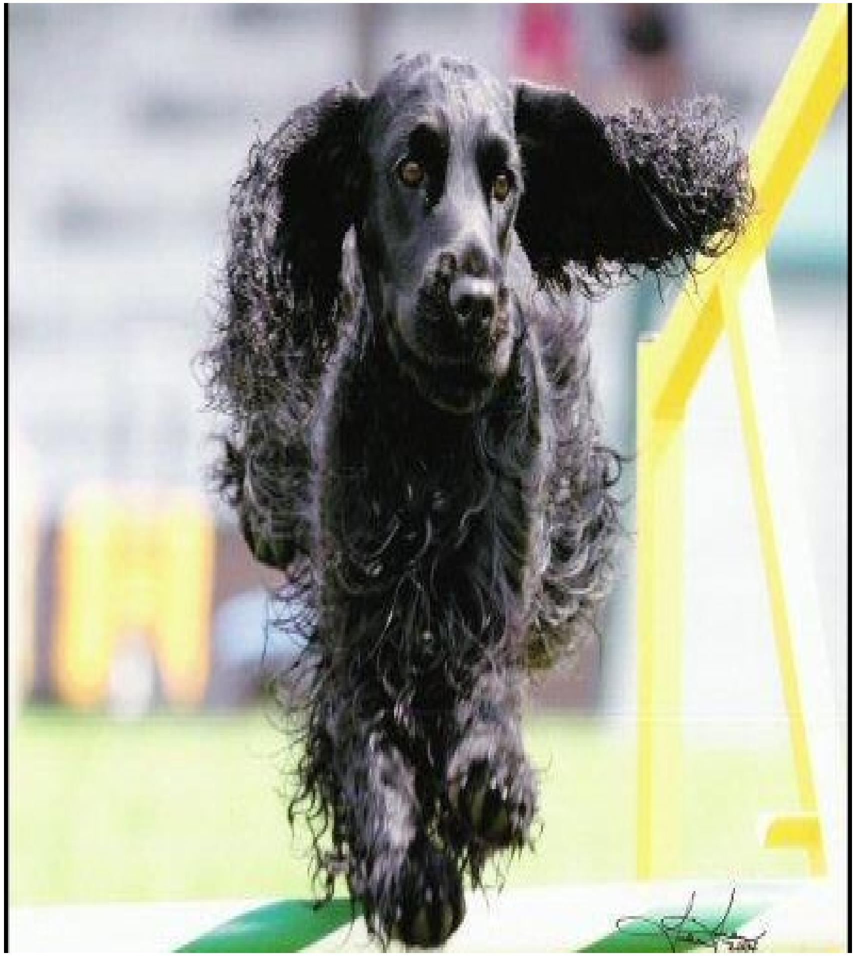
Both labrador retriever



Innumerable Poses

Diverse Appearances

Varying geometry of parts

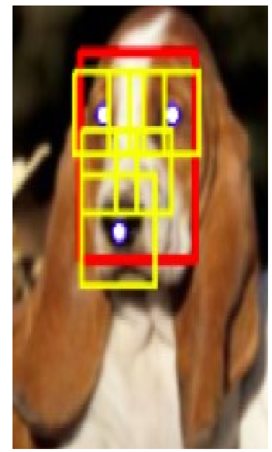


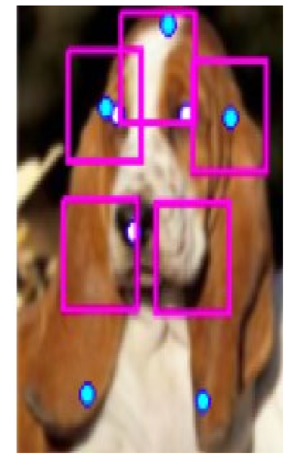
Overview of the system

1. Face Detection

2. Part Detection

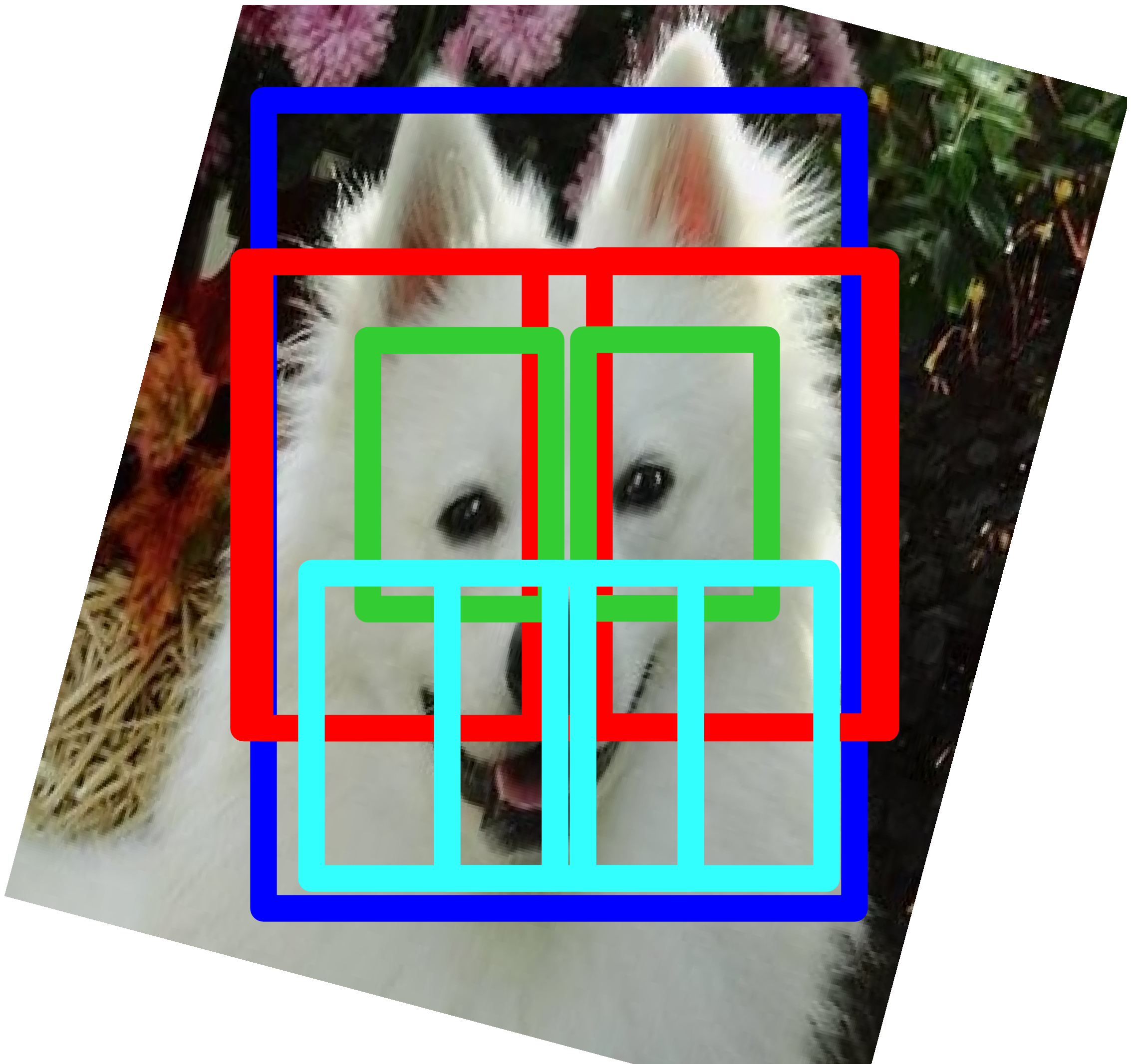
3. Feature Extraction and ear localization





4. One vs All classification

Pipeline 1: Dog Face Detection



Pipeline 2: Localize Parts

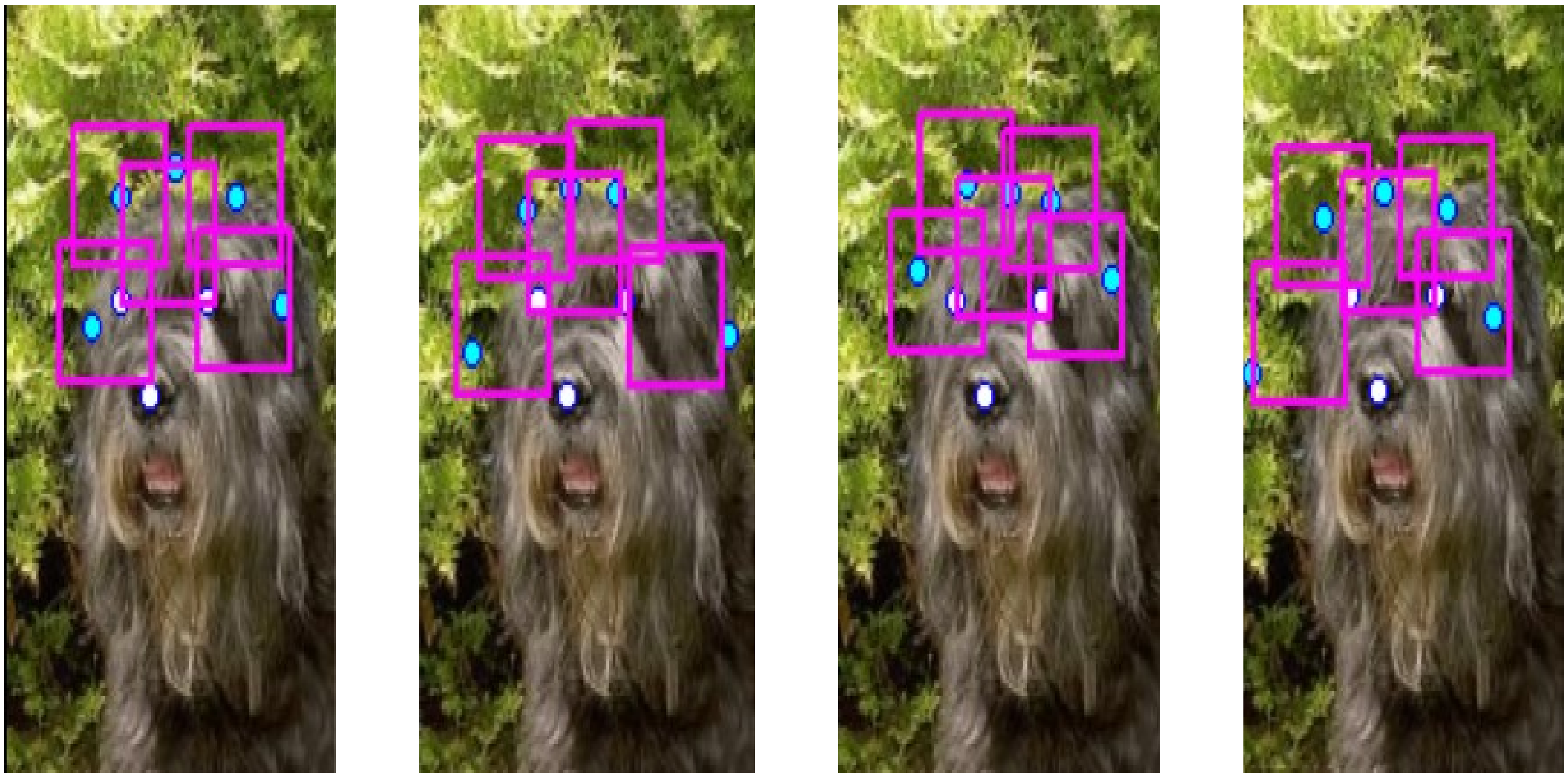
Pipeline 3: Infer ears using detected parts

With *r(=10)* exemplars **from each breed**

Pipeline 3: Infer ears using detected parts

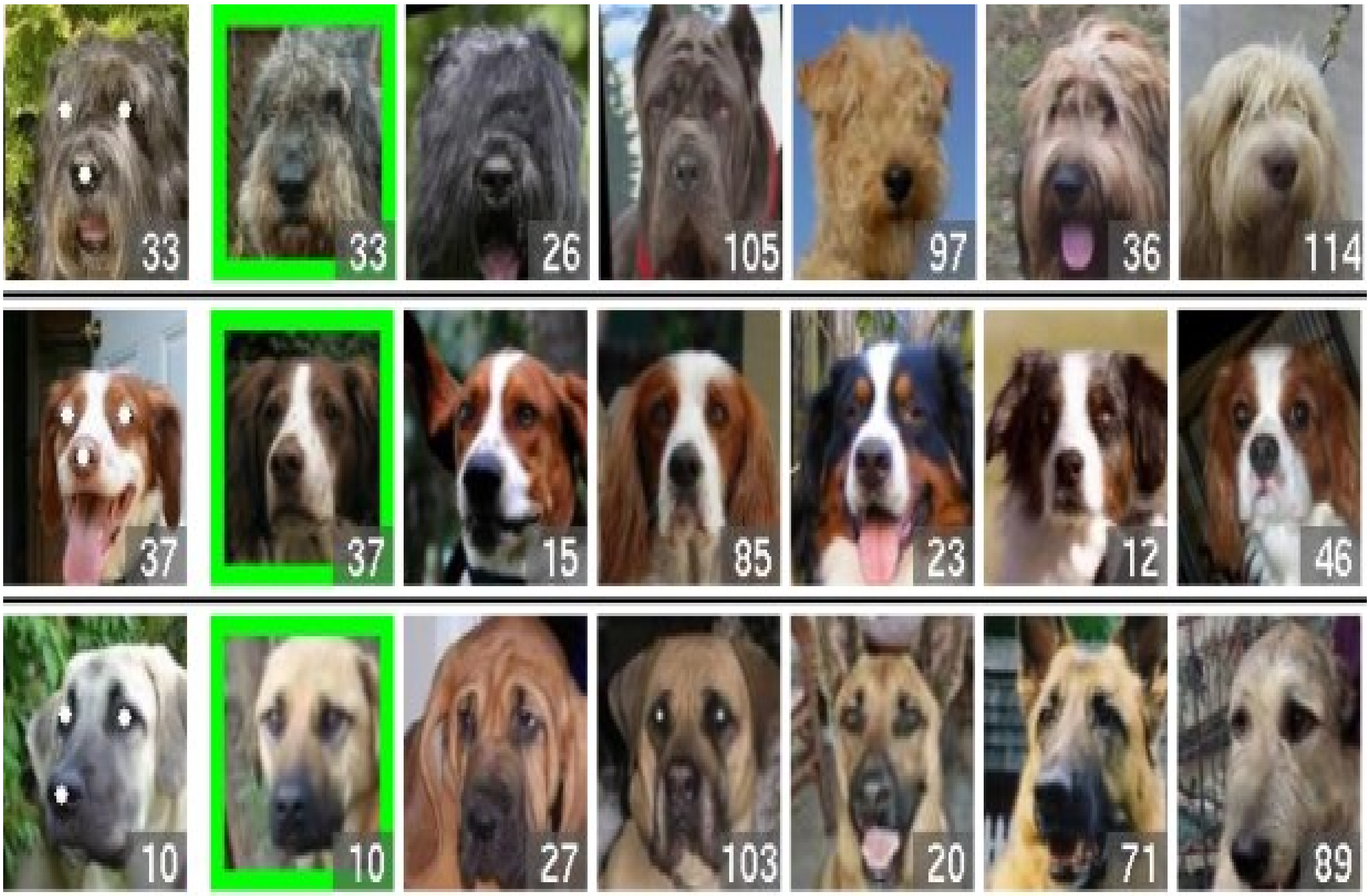
With *r*(=10*)* exemplars from each breed

Pipeline 4: Classification



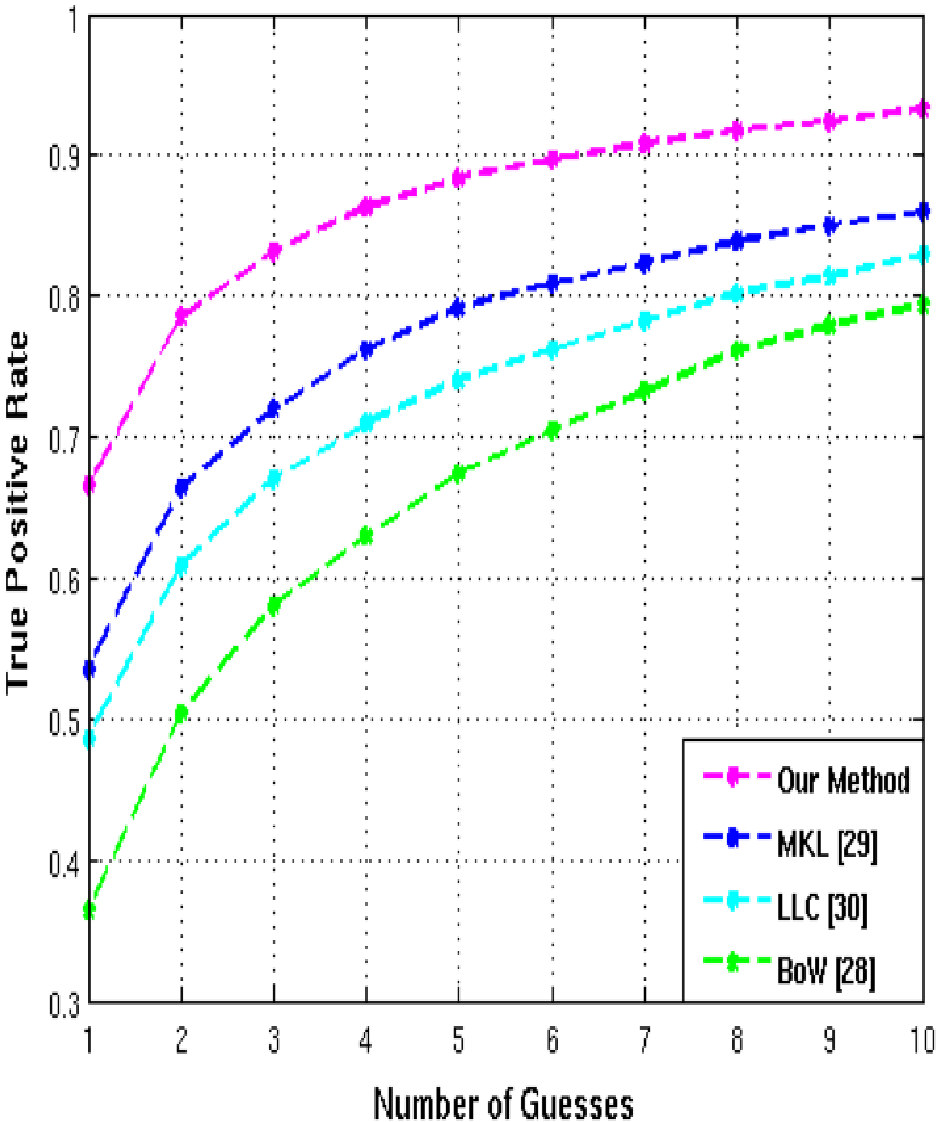
Extract SIFT at part locations for each breed+color histogram → one vs all linear SVM classifier

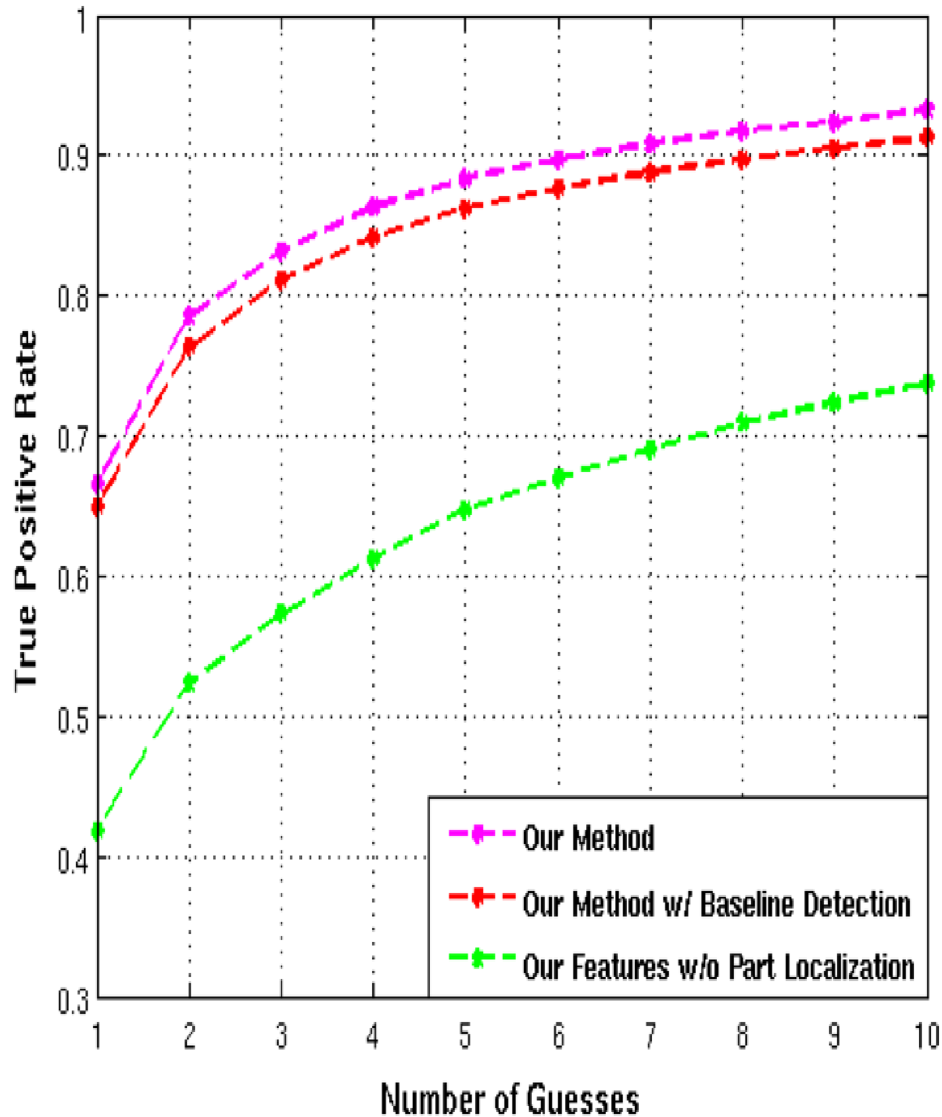
Qualitative Results: Successful



Qualitative Results: Failures

Results: ROC curves





Attributes

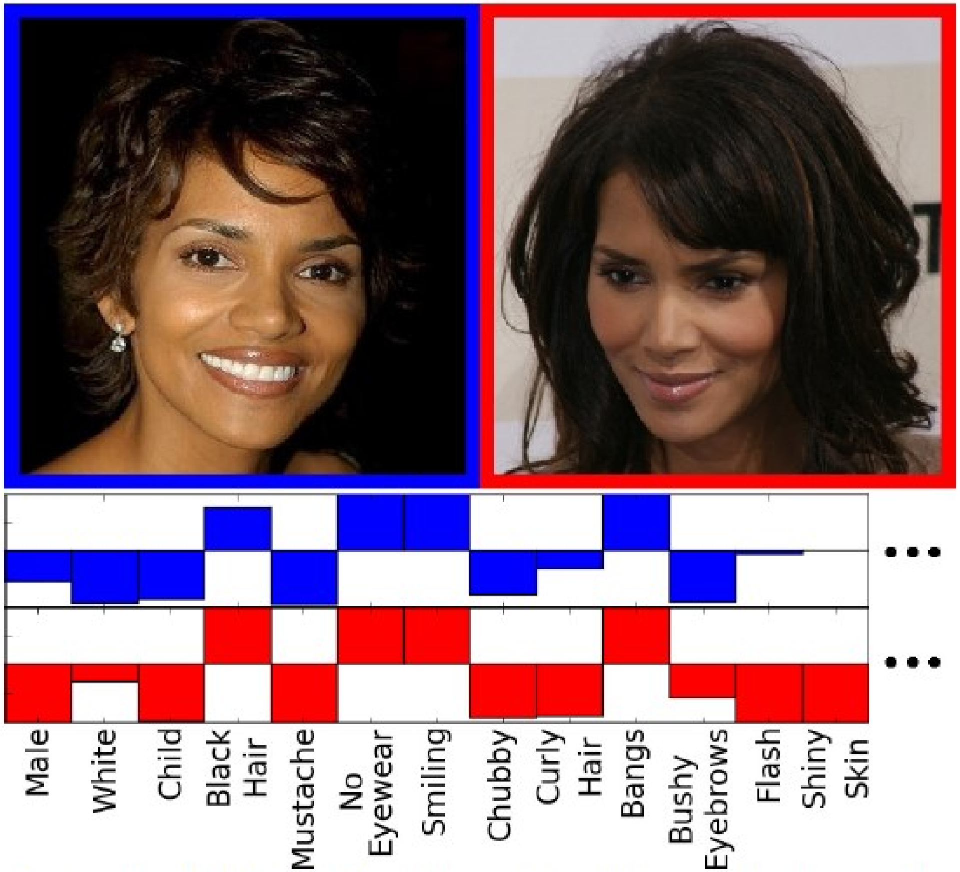
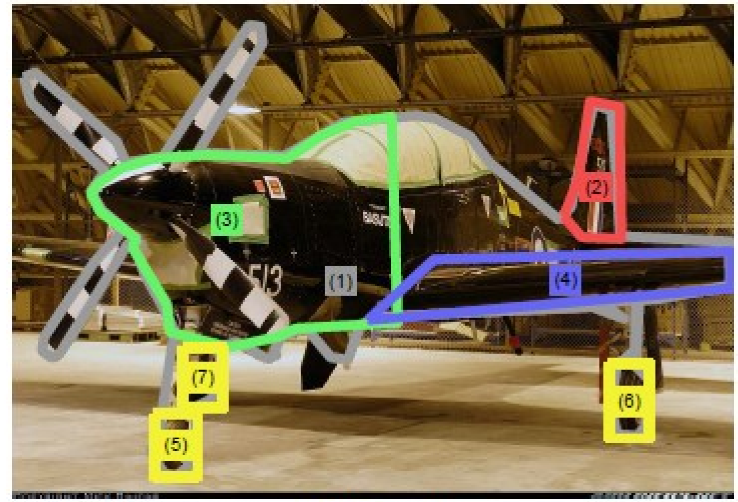


Figure 1: **Attribute Classifiers:** An attribute classifier can be



1 **airplane** facing-direction: SW; is-airliner: no; is-cargo-plane: no;
-glider: no; is-military-plane: yes; is-propellor-plane: yes; is-seaplane:
no; plane-location: on ground/water; plane-size: medium plane;
ıdercarriage-arrangement: one-front-two-back; wing-type: single wing
plane; airline: UK–Air Force; model: Short S-312 Tucano T1
2 **vertical stabilizer** tail-has-engine: no-engine 3 **nose**
s-engine-or-sensor: has-engine 4 **wing** wing-has-engine: no-engine 5
undercarriage cover-type: retractable; group-type: 1-wheel-1-axle;
location: front-middle 6 **undercarriage** cover-type: retractable;
group-type: 1-wheel-1-axle; location: back-left 7 **undercarriage**
ver-type: retractable; group-type: 1-wheel-1-axle; location: back-right