Visipedia: collaborative harvesting and organization of visual knowledge

Pietro Perona
California Institute of Technology

2014 Signal and Image Sciences Workshop
Lawrence Livermore National Laboratory
21 May, 2014
Beak

From Wikipedia, the free encyclopedia

For other uses, see Beak (disambiguation).

The beak, bill or rostrum is an external anatomical structure of birds which, in addition to eating, is used for grooming, manipulating objects, killing prey, probing for food, courtship and feeding their young. The term also refers to a similar mouthpart in some cephalopods, cetaceans, pufferfishes, turtles, Anuran tadpoles and sirens.

Anatomy

Beaks can vary significantly in size and shape from species to species. The beak is composed of an upper jaw called the maxilla, and a lower jaw called the mandible. The jaw is made up of several bones, the most important of which is the maxilla. The maxilla is a thin, flexible bone that allows the beak to be moved. The beak is covered with keratin, the same substance that makes up the skin of our fingernails and hair. The beak may also have a horny sheath, which is a hard, protective layer that covers the beak and is made of keratin. The beak may also have a cartilaginous sheath, which is a soft, flexible layer that covers the beak and is made of cartilage.
The cere (from the Latin cera: wax) or operculum is a soft, fleshy swelling found on the beaks of certain birds. Hawks, parrots, doves, skuas and budgerigars are among the birds that have ceres. The word 'cere' is often used synonymously with 'beak', although the two are not identical. The cere plays a role in indicating the reproductive stage of certain dimorphic birds, and also has a key function in respiration.

Physical characteristics

The cere is located at the top of the beak and is often waxy in texture. The colour of the cere may vary from species to species, and also depends on the season. The cere contains the nares (nostrils). The shape of the cere varies from species to species. In falcons, the opening of the nares is roughly circular in shape.
Can I eat this one? - Love John

Sent from my iPhone
Mushroom

From Wikipedia, the free encyclopedia

For other uses, see Mushroom (disambiguation).

A mushroom is the fleshy, spore-bearing fruiting body of a fungus, typically produced above ground on soil or on its food source. The standard for the name "mushroom" is the cultivated white button mushroom, Agaricus bisporus, hence the word mushroom is most often applied to those fungi (Basidiomycota, Agaricomycetes) that have a stem (stipe), a cap (pileus), and gills (lamellae, sing. lamella) on the underside of the cap, just as do store-bought white mushrooms.

The word "mushroom" can also be used for a wide variety of gilled fungi, with or without stems, and the term is used even more generally, to describe both the fleshy fruiting bodies of some Ascomycota and the woody or leathery fruiting bodies of some Basidiomycota.
Wikipedia says: use a field guide
Yet, the info is there...
Amanita pantherina var. pantherina, also known as the "European Panther" and "False Blusher" due to its similarity to the true Blusher (Amanita rubescens), is a species of Europe and western Asia. Material described as A. pantherina in the Americas seems to belong to a number of distinct taxa only some of which have been described.
Visual expertise: not easily accessible to machines
Femur

From Wikipedia, the free encyclopedia

For the invertebrate femur, see Arthropod leg.

The femur, or thigh bone, is the most proximal (closest to the body) bone of the leg in vertebrates capable of walking or jumping, such as most land mammals, birds, many reptiles such as lizards, and amphibians such as frogs. In vertebrates with four legs such as dogs and horses, the femur is found only in the rear legs.

Contents [hide]
1 Human anatomy
2 Evolutionary variation
3 Etymology
4 Additional images
5 References
6 External links

Human anatomy

In human anatomy, the femur is the longest and largest bone. Along with the temporal bone of the skull, it is one of the two strongest bones in the body. The average adult male femur is 48 centimeters (18.9 in) in length and 2.34 cm (0.92 in) in diameter and can support up to 30 times the weight of an adult.[1] It forms part of the hip (at the acetabulum) and part of the knee, which is located above. There are four eminences, or protuberances, in the human femur: the head, the greater trochanter, the lesser trochanter, and the lower extremity. They appear at various times from just before birth to about age 14. Initially, they are joined to the main body of the femur with cartilage, which gradually ossifies to form bone.
Lessons:

• Visual queries
  • Easy for humans
  • Difficult for machines

• Pictures are *digital dark matter*

• Expert knowledge - how to collect?
Cere

From Wikipedia, the free encyclopedia

The cere (from the Latin cera: wax) are the beaks of certain birds. Hawks, peregrines, geese, and other birds that have cera. The word 'cere' refers to the beak of a bird, and the two are not identical. The cere is one of the key parts of the skull, certain dimorphic birds, and also has a role in respiration.

Contents [hide]
1 Physical characteristics
2 Role in respiration
3 Role in indication of reproductive cycle
4 References
5 See also

Physical characteristics

The cere is located at the top of the bill in texture. The colour of the cere may vary from species to species, and also depends on the age of the bird (the narial). The shape of the cere is species-dependent. In falcons, the opening of the cere is elongated, which helps them to capture their prey.
The *cere* (from the Latin *cera*: wax)\(^1\) or *operculum*\(^2\) is a soft, fleshy swelling found on the beaks of certain birds. Hawks, parrots, doves, skus and budgerigars are among the birds that have cerees. The word 'cere' is often used synonymously with 'beak', although the two are not identical. The cere plays a role in indicating the reproductive stage of certain dimorphic birds, and also has a key function in respiration.

### Contents

1. Physical characteristics
2. Role in respiration
3. Role in indication of reproductive cycle
4. References
5. See also

### Physical characteristics

The cere is located at the top of the beak,\(^2\%^3\) and is often waxy in texture. The colour of the cere may vary from species to species, and also depends on the season.\(^4\) The cere contains the nares (nostrils). The shape of the cere varies from species to species. In falcons, the opening of the nares is roughly circular in shape.\(^5\)
Cere

The cere (from the Latin cera: wax) or operculum is a soft, fleshy swelling found on the beaks of certain birds. Hawks, parrots, doves, skuas and budgerigars are among the birds that have ceres. The word 'cere' is often used synonymously with 'beak', although the two are not identical. The cere plays a role in indicating the reproductive stage of certain dimorphic birds, and also has a key function in respiration.

Physical characteristics

The cere is located at the top of the beak and is often waxy in texture. The colour of the cere may vary from species to species, and also depends on the season. The cere contains the nares (nostrils). The shape of the cere varies from species to species. In falcons, the opening of the nares is roughly circular in shape.
Cere
From Wikipedia, the free encyclopedia

The cere (from the Latin cera: wax) or operculum[1] is a soft, fleshy swelling found on the beaks of certain birds. Hawks, parrots, doves, skus and budgerigars are among the birds that have ceras. The word `cere` is often used synonymously with `beak`, although the two are not identical. The cere plays a role in indicating the reproductive stage of certain dimorphic birds, and also has a key function in respiration.

Contents [hide]
1 Physical characteristics
2 Role in respiration
3 Role in indication of reproductive cycle
4 References
5 See also

Physical characteristics
The cere is located at the top of the beak[2][3] and is often waxy in texture. The colour of the cere may vary from species to species, and also depends on the season.[4] The cere contains the nares (nostrils). The shape of the cere varies from species to species. In falcons, the opening of the nares is roughly circular in shape.[5]
How do we make this happen?
`Visipedia`
Visipedia’

Preprocessing
`Visipedia`

Preprocessing

Expert

- Head
- Eye
- Cere

....
`Visipedia`

Preprocessing

Expert

Head

Eye

Cere

User

....
`Visipedia`

Preprocessing

Expert

User

Head

Eye

Cere

....
Unsupervised learning

[1] Fergus et al., CVPR03
Supervised learning

[Felzenszwalb et al. ’10]
Need for human expertise
Need for human expertise

Throat
Need for human expertise

Throat

- Pharynx
- Epiglottis
- Larynx
- Trachea
If you believe you may have seen an Ivory-billed Woodpecker, immediately after the sighting, make a drawing of what you saw, noting the following characteristics:

- Color of trailing edge of wing (white vs. black)
- Crest and forehead color
- Bill color
- Chin color
- Relative size
- Vocalizations
- Habitat

If you are with someone else, individuals should make their own notes without conferring with each other.

Please report any sightings to refuge officials and to the Cornell Lab of Ornithology via its website: http://www.birds.cornell.edu/ivory

Identifying Field Marks of an Ivory-billed Woodpecker and Similar Birds

**In flight - view from below**

*Distinct Ivory-billed Woodpecker characteristics:*
- White trailing edge of wing (vs. dark trailing edge of Pileated).
- Wing more slender than Pileated.
- Tail feathers longer and more pointed.
- Pale, ivory-white bill.

**In flight - view from above**

*Distinct Ivory-billed Woodpecker characteristics:*
- White trailing edge of wing (vs. dark trailing edge of Pileated).
- Two white stripes converge on lower back.
- Tail feathers longer and more pointed.
- Pale, ivory-white bill.

**At rest**

*Distinct Ivory-billed Woodpecker characteristics:*
- Two white stripes converge on lower back.
- Entirely white secondary feathers give appearance of white “saddle” on back.
- Largely dark face and dark chin (vs. white chin of Pileated).
- Pale, ivory-white bill.
- Crest is curved and pointed; male crest is red with black forehead (Pileated male crest is entirely red).

*Female Head*
- Female Ivory-bill crest is entirely black (female Pileated crest resembles male ivory-billed red crest with black forehead – use chin color as distinguishing feature)
THE PARTS OF A LOCOMOTIVE

- cab
- number board
- exhaust stack
- dynamic-brake air intake
- dynamic-brake cooling fans
- radiator cooling far
- radiator air intakes
- rear platform
- engine access doors
- air reservoir
- battery box
- frame or platform
- fuel tank
- equipment blower duct
- 3-axle C-C truck
- nose or low hood
- pilot

Note:
Some handrails are omitted for clarity
WIKIPEDIA

VISIPEDIA
Images, segments, annotations, links, GUIs, diagnostics

Users
VISIPEDIA
Images, segments
annotations, links,
GUs, diagnostics

Users

Image databases

WIKIPEDIA

flickr

IMAGE NET
VISIPEDIA

Images, segments, annotations, links, GUIs, diagnostics

WIKIPEDIA

Users

Experts

Annotators

Automata

Image databases

flickr

IMAGENET

amazon mechanical turk
VISIPEDIA

Images, segments, annotations, links, GUIs, diagnostics

Users

Experts

Annotators

Automata

Vision scientists

Image databases

WIKIPEDIA

flickr

IMAGENET
Crowdsourcing image annotation

[Welinder et al., NIPS2010]
Indigo Bunting
5,926 images
Building datasets

6000 images from flickr.com

Filter

100s of training images
Building datasets

Annotators

Is there an Indigo bunting in the image?

6000 images from flickr.com

100s of training images
Find the Indigo Bunting
Find the Indigo Bunting
Characterizing annotators: types of errors

<table>
<thead>
<tr>
<th>Indigo Bunting?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Alarm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Rejection</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Task: Find the Indigo Bunting

hit rate (correct detection) vs. rate of correct rejection
Task: Find the Indigo Bunting

Rate of correct rejection vs. hit rate (correct detection)

- 6% error
- 15% error
- 31% error
- 50% error
Task: Find the Indigo Bunting

Rate of correct rejection vs. hit rate (correct detection) graph.
Task: Find the Indigo Bunting

Rate of correct rejection

Hit rate (correct detection)

6% error

15% error

31% error

50% error

Bots
Task: Find the Indigo Bunting

optimists

rate of correct rejection

hit rate (correct detection)

6% error

15% error

31% error

50% error
Task: Find the Indigo Bunting

- 50% error
- 31% error
- 15% error
- 6% error

pessimists
Task: Find the Indigo Bunting

- 50% error
- 31% error
- 15% error
- 6% error

hit rate (correct detection)

rate of correct rejection

adversaries
Image difficulty and annotator competence

$\text{image } i \\ \rightarrow \mathcal{Z}_i \rightarrow \text{object presence} \\ \downarrow \\
\text{viewpoint} \\ \text{specimen} \\ \text{location} \\ \text{weather} \\ \text{camera} \\ \text{pose} \\ \downarrow \\
\mathcal{I}_i \rightarrow \text{image}$
Image difficulty and annotator competence

- Image $i$
- Object presence $Z_i$
- Neural processing $I_i$
- Signal seen by ideal observer $X_i$
- Viewpoint
- Specimen
- Location
- Weather
- Camera
- Pose

$p(x_i | z_i = 0)$
$p(x_i | z_i = 1)$
Image difficulty and annotator competence

$p(x_i | z_i = 0)$

$z_i \rightarrow I_i \rightarrow x_i$

image $i$

viewpoint
specimen
location
weather
camera
pose

signal seen
by ideal observer

$-3 -2 -1 0 1 2 3$
Image difficulty and annotator competence

image $i$

$z_i$

$\mathcal{I}_i$

$\mathcal{X}_i$

viewpoint
specimen
location
weather
camera
pose

signal seen by ideal observer

$p(x_i \mid z_i = 0)$

$x_1$

$-3 -2 -1 0 1 2 3$

$x_i$

Image difficulty and annotator competence

$z_i$

$\mathcal{I}_i$

$\mathcal{X}_i$

viewpoint
specimen
location
weather
camera
pose

signal seen by ideal observer

$p(x_i \mid z_i = 0)$

$x_1$

$-3 -2 -1 0 1 2 3$

$x_i$
Image difficulty and annotator competence

image $i$

$z_i$

$I_i$

$X_i$

viewpoint
specimen
location
weather
camera
pose

signal seen
by ideal observer

$p(x_i | z_i = 0)$

$x_1$

$X_i$

$-3 -2 -1 0 1 2 3$
Image difficulty and annotator competence

image $i$

$z_i$

viewpoint
specimen
location
weather
camera
pose

$I_i$

signal seen by ideal observer

$x_i$

$\sigma_j$

annotator noise

signal annotator $j$ sees

$\mathbb{P}(x_i \mid z_i = 0)$

$X_1$

$-3$ $-2$ $-1$ $0$ $1$ $2$ $3$ $x_i$
Image difficulty and annotator competence

image $i$

$z_i$

viewpoint
specimen
location
weather
camera
pose

$I_i$

signal seen
by ideal observer

$x_i$

$\sigma_1 = 0.2$

$\sigma_j$

signal annotator $j$ sees

$y_{ij}$

$p(x_i \mid z_i = 0)$

$x_1$

$\sigma_1 = 0$

$y_{i1}$
Image difficulty and annotator competence

- Image $i$
- Viewpoint
- Specimen location
- Weather
- Camera pose

- Signal seen by ideal observer $x_i$
  - $\sigma_1 = 0.2$

- Annotator noise $\sigma_j$
- Signal annotator $j$ sees $y_{ij}$
  - $\sigma_2 = 0.5$

$p(x_i \mid z_i = 0)$

- Image $i$:
  - $x_1$

- Annotator $j$:
  - $y_{i1}$
  - $y_{i2}$
Image difficulty and annotator competence

- **image** $i$
- **viewpoint**
- **specimen location**
- **weather**
- **camera pose**

$z_i$ → $I_i$ → $x_i$ → $y_{ij}$ → $\sigma_j$

signal seen by ideal observer

annotator noise

$p(x_i | z_i = 0)$

- $\sigma_1 = 0.2$
- $\sigma_2 = 0.5$
- $\sigma_3 = 5.0$

$y_{i1}$ → $y_{i2}$ → $y_{i3}$

Signal seen by ideal observer
**Image difficulty and annotator competence**

- **image** $i$
- **viewpoint** $z_i$
- **specimen location** $I_i$
- **weather** $x_i$
- **camera pose** $\sigma_j$
- **annotator noise** $Y_{ij}$
- **signal annotator $j$ sees** $p(x_i \mid z_i = 1)$

$$x_i$$
Image difficulty and annotator competence

- Image $i$
- Viewpoint
- Specimen location
- Weather
- Camera pose

- $Z_i$
- $I_i$
- $x_i$

- Annotator noise $\sigma_j$
- Signal annotator $j$ sees

$p(x_i \mid z_i = 1)$

$\mathbf{X}_i$
Image difficulty and annotator competence

- Image $i$
  - viewpoint
  - specimen location
  - weather
  - camera pose
- $z_i$
- $I_i$
- $x_i$
- Annotator noise $\sigma_j$
- Signal annotator $j$ sees

$$p(x_i \mid z_i = 1)$$

$x_i$ distribution: $-3, -2, -1, 0, 1, 2, 3$
Image difficulty and annotator competence

image $i$

$z_i$

viewpoint
specimen
location
weather
camera
pose

$I_i$

$p(x_i \mid z_i = 1)$

$\sigma_j$

$Y_{ij}$

annotator noise

signal annotator $j$ sees

$\mathbf{x}_i$
Image difficulty and annotator competence

image $i$

$z_i$

viewpoint, specimen location, weather, camera pose

$I_i$

$\mathbb{P}(x_i | z_i = 1)$

$\sigma_j$

annotator noise

$y_{ij}$

signal annotator $j$ sees

$\mathcal{X}_i$
Image difficulty and annotator competence

image $i$

$z_i$

$\sigma_j$

$\sigma_1 = 0.2$

$\sigma_2 = 0.5$

$\sigma_3 = 5.0$

$p(x_i \mid z_i = 1)$

viewpoint
specimen location
weather camera pose

annotator noise

signal annotator $j$ sees

$x_i$

$y_{i1}$

$y_{i2}$

$y_{i3}$
Optimists and pessimists: $\tau_j$

\[
p(y_{ij} \mid z_i = 0) \quad p(y_{ij} \mid z_i = 1)
\]
Optimists and pessimists: $\mathcal{T}_j$

$p(y_{ij} \mid z_i = 0)$

$p(y_{ij} \mid z_i = 1)$

$\mathcal{T}_j$

label $l_{ij} = 1$
Optimists and pessimists: $\tau_j$

$p(y_{ij} | z_i = 0)$

$p(y_{ij} | z_i = 1)$

Label $l_{ij} = 0$
p(y_{ij} \mid z_i = 0) \quad p(y_{ij} \mid z_i = 1)

\tau_j

Optimists and pessimists: $\tau_j$
Optimists and pessimists: $\tau_j$

$p(y_{ij} \mid z_i = 0)$

$p(y_{ij} \mid z_i = 1)$

hit rate (correct detection)

rate of correct rejection

pessimist
Optimists and pessimists: $\tau_j$

$p(y_{ij} \mid z_i = 0)$  $p(y_{ij} \mid z_i = 1)$

hit rate (correct detection)
Optimists and pessimists: $\tau_j$

$$p(y_{ij} \mid z_i = 0)$$  $$p(y_{ij} \mid z_i = 1)$$

$\tau_j$
Multidimensional signals and annotators

\[ \mathbf{x}_i = (x^1_i, x^2_i) \]

\[ p(\mathbf{x}_i \mid z_i = 1) \]

\[ p(\mathbf{x}_i \mid z_i = 0) \]
Multidimensional signals and annotators

\[ x_i = (x_i^1, x_i^2) \]

\[ p(x_i \mid z_i = 1) \]

\[ p(x_i \mid z_i = 0) \]
Multidimensional signals and annotators

\[ \mathbf{x}_i = (x^1_i, x^2_i) \]

\[ p(\mathbf{x}_i | z_i = 1) \]

\[ p(\mathbf{x}_i | z_i = 0) \]

\[ \mathbf{w}_j = (w^1_j, w^2_j) \]

\[ \tau_j \]
Full model

\[ x_i \rightarrow \{\beta, \theta_z\} \rightarrow z_i \rightarrow \{\mathcal{I}_i, N\} \rightarrow y_{ij} \rightarrow l_{ij} \rightarrow |\mathcal{L}_{ij}| \]

\[ \alpha \rightarrow \sigma_j \rightarrow w_j \rightarrow \tau_j \rightarrow M \]

images

annotators

labels
Full model

- Images: $Z_i$ → $x_i$
- Parameters: $\beta$, $\theta_z$
- Labels: $\mathcal{L}_{ij}$
- Annotators: $\alpha$, $\gamma$, $\sigma_j$, $w_j$, $\tau_j$
- $N$ images, $M$ annotators, $|\mathcal{L}_{ij}|$ labels
Full model

images

Zi

\[ \beta \]

zi

xi

\[ \theta_z \]

\[ N \]

\[ i \]

\[ J_i \]

\[ \mathcal{L}_{ij} \]

\[ \alpha \]

\[ \gamma \]

\[ \sigma_j \]

\[ w_j \]

\[ \tau_j \]

\[ M \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]

\[ \alpha \]

\[ \gamma \]
Full model

\[\begin{align*}
\beta & \rightarrow Z_i \\
& \rightarrow x_i \\
& \rightarrow y_{ij} \\
\theta_{z} & \rightarrow y_{ij} \\
\end{align*}\]

images \( N \) 

\[\begin{align*}
\sigma_j & \rightarrow y_{ij} \\
\tau_j & \rightarrow y_{ij} \\
\end{align*}\]

annotators \( M \) 

\[\begin{align*}
l_{ij} & \rightarrow \mathcal{L}_{ij} \end{align*}\]
Full model

\[
\begin{align*}
\beta & \quad \theta_z \\
Z_i & \quad x_i \\
\mathcal{I}_i & \\
\end{align*}
\]

\[
\begin{align*}
\sigma_j & \quad w_j \\
j & \quad \tau_j \\
\end{align*}
\]

\[
\begin{align*}
y_{ij} & \quad l_{ij} \\
i & \quad |\mathcal{L}_{ij}| \\
L & \\
\end{align*}
\]

images

annoctors

labels
Full model

\[ z_i \xrightarrow{\beta} x_i \xrightarrow{\theta_z} y_{ij} \xrightarrow{\alpha} \sigma_j \xrightarrow{\gamma} l_{ij} \]

Images: \( z_i \)

\( i \) \( j \)

Labels: \( l_{ij} \)

Annotators: \( \sigma_j \), \( w_j \), \( \tau_j \)

\( |L_{ij}| \)

\( M \)
Full model

\[ \beta \quad \theta_z \]

Images

\[ Z_i \rightarrow x_i \]

Labels

\[ y_{ij} \rightarrow l_{ij} \]

Annotators

\[ \sigma_j \quad w_j \quad \tau_j \]

\[ \alpha \quad \gamma \]

\[ |L_{ij}| \]
Experiments: estimating $z_i$

Indigo Bunting

Blue Grosbeak
Experiments: estimating $\gamma_i$

![Graph showing error rate vs number of annotators]

- **Our model**
- **Majority**
- **Bias**
- **Glad**

Dawid & Skene '79
Whitehill et al. '09
Experiments: estimating $\tilde{z}_i$

- **our model**
- **majority**
- **bias**
- **glad**

Dawid & Skene '79
Whitehill et al. '09

Error rate vs. number of annotators

- 20% error
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?

Annotators
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?

Annotations

Many binary labels
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?

Annotators

Many binary labels

Model + Inference
Sanity Check 1: Rotated Ellipses Experiment

Is the ellipse oriented vertically?

Annotators

Many binary labels

Model + Inference

image difficulty

estimated $x_i$

deviceation (degrees)
Sanity Check 2: Rotated Ellipses Experiment

$\tau_j$

$\frac{1}{\sigma_j}$
Greebles: estimating multiple attributes

Synthetic figures distinguished by height and color

\[ x_i = (x_i^1, x_i^2) \]

yellow

tall

short

green

short

orange
Greebles: estimating multiple attributes

Send to two groups of annotators

1. Select green greebles

2. Select tall greebles
Greebles: estimating multiple attributes
Greebles: estimating multiple attributes

\[ x_i^2 \]

\[ x_i^1 \]
Greebles: estimating multiple attributes
Greebles: estimating multiple attributes
Greebles: estimating multiple attributes

\[
x_i = (x_i^1, x_i^2)
\]
Greebles: estimating multiple attributes
Greebles: estimating multiple attributes

Select tall greebles

Select green greebles
Last Experiment: Waterbirds

Mallard

American Black Duck

Canada Goose

Red Necked Grebe

Non-bird
Waterbirds

Mallard  American Black Duck

Canada Goose  Red Necked Grebe  Non-bird
Is there a duck in the image?
Is there a duck in the image?
Is there a duck in the image?
Is there a duck in the image?
Is there a duck in the image?
Is there a duck in the image?
Is there a duck in the image?
Is there a duck in the image?
The first application: sleep annotation

[Warby et al. 2014]
[Warby et al. 2014]
Supplementary Figure 1A: The trade-off between precision and recall for different models. Each point represents a model's performance, with the line connecting these points indicating the model's precision-recall curve. The models are labeled as a2, a4, a5, a6, a1, and a3.
Adaptive on-line annotation

H vs T what is the bias?
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: \( H \) vs \( T \) what is the bias?
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: $\text{H} \quad \text{vs} \quad \text{T}$ what is the bias?
Adaptive on-line annotation

- How many labels do we need until we are certain of an annotation?

- Coin flipping: \( \text{H} \) vs \( \text{T} \) what is the bias?

  \( \text{T} \text{T} \)
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: $H$ vs $T$ what is the bias?

  $T$ $T$ $H$
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping:  H vs T what is the bias?

  T  T  H  H
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: \( H \) vs \( T \) what is the bias?

\[ T \ T \ H \ H \ H \]
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: H vs T what is the bias?

T T H H H T
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: $H$ vs $T$ what is the bias?
Adaptive on-line annotation

- How many labels do we need until we are certain of an annotation?
- Coin flipping: $\text{H}$ vs $\text{T}$ what is the bias?

\begin{figure}[h]
\centering
\begin{tabular}{cccccccc}
T & T & H & H & H & T & T & T \\
\end{tabular}
\end{figure}
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: \( \text{H} \) vs \( \text{T} \) what is the bias?

\[
\begin{array}{cccccccccc}
\text{T} & \text{T} & \text{H} & \text{H} & \text{H} & \text{T} & \text{T} & \text{T} & \text{T} \\
\end{array}
\]
Adaptive on-line annotation

- How many labels do we need until we are certain of an annotation?
- Coin flipping: $H$ vs $T$ what is the bias?

$T$ $T$ $H$ $H$ $H$ $T$ $T$ $T$ $T$ $T$
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: \( H \) vs \( T \) what is the bias?

• Wald (1945): sequential probability ratio test (SPRT)
Adaptive on-line annotation

• How many labels do we need until we are certain of an annotation?

• Coin flipping: H vs T what is the bias?

• Wald (1945): sequential probability ratio test (SPRT)
Adaptive on-line annotation

- How many labels do we need until we are certain of an annotation?

- Coin flipping: \( H \) vs \( T \) what is the bias?

- Wald (1945): sequential probability ratio test (SPRT)
We have proposed an online algorithm to determine with the uncertainty of their estimated ground truth values. We actively select which images to label based on the probability the 'ground truth value' of some property is high. Our algorithm can estimate this quantity as estimated by the algorithm decreases counting errors only for images where the full model using all labels per image is large. Notice in the figure that is worth noting that most of the errors made by the online algorithm are on images where the intrinsic uncertainty of the ground truth label is high. We have shown how the algorithm can be applied to different types of annotations commonly used in computer vision. We also find that equally skilled annotators differ in the reliability of the product it produces an estimate of annotator expertise and prioritizing experts online. By combining just the right number of noisy annotations it defines an optimal 'virtual expert' consistently providing high quality labels. We have also shown that there are differences in the number of noisy annotations needed for the batch algorithm to achieve the same performance as the online algorithm with limited number of labels per image. For the Presence dataset twice as many labels per image are needed for the batch algorithm to achieve the same performance as the online algorithm. With the Presence dataset the uncertainty in the ground truth labels is high. The online algorithm are on images where the intrinsic uncertainty of the ground truth label is high. This means that almost three times lower than the general algorithm when the task is difficult. Through it will require few labels per image. 

8. Conclusions

To get an idea of the performance of the online algorithm we compared it to running the batch version from the uncertainty of their estimated ground truth values. We have shown how the algorithm can be applied to different types of annotations commonly used in computer vision. We have also shown that there are differences in the number of noisy annotations needed for the batch algorithm to achieve the same performance as the online algorithm with limited number of labels per image. For the Presence dataset twice as many labels per image are needed for the batch algorithm to achieve the same performance as the online algorithm. With the Presence dataset the uncertainty in the ground truth labels is high. This means that almost three times lower than the general algorithm when the task is difficult. Through it will require few labels per image. 

CVPR 2010 Submission #7. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.
Crowdclustering

[Gomes et al., NIPS 2011]
Organizing visual knowledge
Clustering
Need

$x_i \in \mathbb{R}^d$
Need

\[ A_{ij} \in (0, 1) \]
Cluster this...
Metric for clustering
Showing 1 - 18 of 67,880 Results, sorted by `Price: High-to-Low` [3 June 2011]
Challenges

- Lots of images
- Annotators working on small subsets
- Categorization criteria may differ
Generative Model

- Assume images drawn from clusters in an embedding space
- Each annotator corresponds to an inner product in that space
How do we aggregate the results from the crowd?

- Cluster 1
- Cluster 2

- Annotator 1: sensitive to ground vs. air
How do we aggregate the results from the crowd?

• Annotator 2: sensitive to left vs right
Generative Model
Generative Model

• Annotator 1: sensitive to ground vs. air
• Annotator 2: sensitive to left vs. right
\[ p(l_{ab2} = 1 | \mathbf{x}_a, \mathbf{x}_b, \mathbf{W}_2, \tau_2) = \frac{1}{1 + \exp\left\{ - (\mathbf{x}_a^T \mathbf{W}_2 \mathbf{x}_b - \tau_2) \right\} } \]
Model

Data Items

"Platonic" clusters

Annotators

Pairwise Labels

\[ Z_i \] 

\[ V_k \]

\[ X_i \]

\[ \Phi_k \]

\[ W_j \]

\[ \tau_j \]

\[ \text{l}_{\text{abj}} \]

\[ \text{abj} \]
Pipeline

Image set

Annotators

Pairwise labels

Model / inference

Categories

"Platonic" clusters

\[ \vdash = 6 \quad \vdash = 6 \quad \vdash = 6 \]

\[ W_i \]

\[ \tau_j \]

\[ 1_t \]

\[ \text{Annotators} \]

\[ \text{Pairwise Labels} \]

\[ \text{Data Items} \]

\[ \Phi_k \]

\[ Z_i \]

\[ V_k \]
Experiments
Average assignment entropy (bits): 0.0029653
<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>bedroom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suburb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kitchen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>living room</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>highway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inside city</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mountain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>open country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>street</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tall building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>office</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inferred Cluster
Average assignment entropy (bits): 0.004792
Crowdclustering

- Experiments \(\sim 10^4\) images, \(\sim 10^2\) workers
- Works well
- Better than one expert
- Any discoveries?
Summary

- Vision and Knowledge
- Visipedia
- Crowdsourcing visual processing
- Crowdclustering

http://www.vision.caltech.edu/visipedia/
Collaborators

• Serge Belongie
• Steve Branson
• Catherine Wah
• Peter Welinder
• Ryan Gomes

http://www.vision.caltech.edu/visipedia/