Understanding Classification Decisions for Object Detection

Will Landecker, Mick Thomure
PSU Computer Science Graduate Program

Melanie Mitchell
PSU Computer Science Faculty
Classifier
Did the computer vision system make the right decision for the *right reason*?

- Interpreting experimental results
- Making better computer vision systems
- Achieving “human level” computer vision
Outline

• Object detection
  ▫ Ambiguity in measure of accuracy
• Some approaches that address this ambiguity
  ▫ Better data
  ▫ Probabilistic (Bayesian)
  ▫ Object localization
• Our approach
  ▫ Visualization
  ▫ Relevant Accuracy: new measure of classification
Outline

• Object detection
  ▫ Ambiguity in measure of accuracy
• Some approaches that address this ambiguity
  ▫ Better data
  ▫ Probabilistic (Bayesian)
  ▫ Object localization
• Our approach
  ▫ Visualization
  ▫ Relevant Accuracy: new measure of classification
Object detection

- Deciding whether or not a specified object exists in an image

- Natural images are hard

- Necessary but not sufficient for “human-level” vision
Object detection

Manually separate images into **positive** (object is present) and **negative** (object is absent)
Object detection
### Object detection

<table>
<thead>
<tr>
<th></th>
<th>0.1</th>
<th>0.5</th>
<th>0.3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.9</th>
<th>0.7</th>
<th>0.1</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.2</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>0.1</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

- Train a classifier (e.g. naïve Bayes, SVM, neural network)
- Test with new images

1. New Image
2. Extract Features
3. Classify features
4. Classifier

- Positive (animal)
- Or
- Negative (no animal)
Object detection

• **Accuracy** = \[\frac{\text{# correct classifications}}{\text{# total classifications}}\]

  ▫ **Accuracy alone isn’t enough (tanks)**

• Were the right decisions made for the *right reason*?
  ▫ **Object vs. background**
  ▫ **Spurious statistics in data**
Outline

• Object detection
  ▫ Ambiguity in measure of accuracy
• Some approaches that address this ambiguity
  ▫ Better data
  ▫ Probabilistic (Bayesian)
  ▫ Object localization
• Our approach
  ▫ Visualization
  ▫ *Relevant Accuracy:* new measure of classification
Outline

• Object detection
  ▫ Ambiguity in measure of accuracy
• Some approaches that address this ambiguity
  ▫ Better data
  ▫ Probabilistic (Bayesian)
  ▫ Object localization
• Our approach
  ▫ Visualization
  ▫ *Relevant Accuracy*: new measure of classification
Our approach

Was the right decision made for the right reason?
Our approach

Was the right decision made for the right reason?

What was the reason?

Visualization
Our approach

Was the right decision made for the right reason?

What was the reason?
Visualization

Was the reason correct?
Relevant accuracy of classification
Visualization

- Natural way to understand what pixels caused the classification
Visualization

• Natural way to understand what pixels caused the classification
• Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

Given a feature vector \( \bar{x} = (x_1, x_2, \ldots, x_n) \), we classify with the binary classification function

\[
\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]
\]
Visualization

• Natural way to understand what pixels caused the classification
• Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

Given a feature vector \( \bar{x} = (x_1, x_2, ..., x_n) \), we classify with the binary classification function

\[
\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]
\]

E.g. for naïve Bayes, \( f(x_i) = \ln \frac{P(x_i|\omega^+)}{P(x_i|\omega^-)} \)

and for an unbiased linear kernel SVM, \( f(x_i) = \sum_{s \in \text{S.V.}} c_s \cdot x_{si} x_i \)
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

Given a feature vector $\bar{x} = (x_1, x_2, ..., x_n)$, we classify with the binary classification function

$$\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]$$

Thus the contribution of feature $x_i$ is simply $f(x_i)$
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features *contributed* more to the classification than others
  II. Features are extracted from a specific region of the image
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

The “standard model” of primate visual cortex (Fukushima, 1980; Serre, Oliva and Poggio, 2007)

The cells in every other layer are performing a max operation over a local spatial neighborhood.

Neocognition image from http://ecalifornian.com/image/neocognitron.png
Visualization

• Natural way to understand what pixels caused the classification
• Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

The “standard model” of primate visual cortex (Fukushima, 1980; Serre, Oliva and Poggio, 2007)

\[ y = \max_i x_i \]

The cells in every other layer are performing a max operation over a local spatial neighborhood.
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

The “standard model” of primate visual cortex (Fukushima, 1980; Serre, Oliva and Poggio, 2007)

\[ y = \max_i x_i \]

The cells in every other layer are performing a max operation over a local spatial neighborhood.
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image

The “standard model” of primate visual cortex (Fukushima, 1980; Serre, Oliva and Poggio, 2007)

The cells in every other layer are performing a max operation over a local spatial neighborhood.

\[ y = \max_i x_i \]
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image
- Making a visualization of $\hat{y}(x_1, x_2, \ldots, x_n)$:

  For each $i$, draw $f(x_i)$ on the image pixels from which $x_i$ was extracted

Remember that

$$\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]$$
Visualization

• Natural way to understand what pixels caused the classification
• Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image
• Making a visualization of $\hat{y}(x_1, x_2, \ldots, x_n)$:

For each $i$, draw $f(x_i)$ on the image pixels from which $x_i$ was extracted

Remember that

$$\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]$$
Visualization

• Natural way to understand what pixels caused the classification
• Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image
• Making a visualization of $\hat{y}(x_1, x_2, \ldots, x_n)$:

For each $i$, draw $f(x_i)$ on the image pixels from which $x_i$ was extracted

Remember that

$$\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]$$
Visualization

- Natural way to understand what pixels caused the classification
- Conditions for visualizing:
  I. Some features contributed more to the classification than others
  II. Features are extracted from a specific region of the image
- Making a visualization of \( \hat{y}(x_1, x_2, \ldots, x_n) \):

For each \( i \), draw \( f(x_i) \) on the image pixels from which \( x_i \) was extracted

\[
\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]
\]
Visualization

• Natural way to understand what pixels caused the classification
• Conditions for visualizing:
  I. Some features *contributed* more to the classification than others
  II. Features are extracted from a specific region of the image
• Making a visualization of $\hat{y}(x_1, x_2, \ldots, x_n)$:

For each $i$, draw $f(x_i)$ on the image pixels from which $x_i$ was extracted

Remember that

$$\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]$$
A new measure of classification

How was the object classified?

\[ \hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right] \]

\[ = \text{sgn} \left[ \sum_{i \in \mathcal{O}} f(x_i) + \sum_{j \in \mathcal{B}} f(x_j) \right] \]
A new measure of classification

How was the object classified?

\[ \hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right] \]

\[ = \text{sgn} \left[ \sum_{i \in A} f(x_i) + \sum_{j \in B} f(x_j) \right] \]
A new measure of classification

How was the object classified?

\[ \hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right] \]

\[ \hat{y}_O := \text{sgn} \left[ \sum_{i \in O} f(x_i) \right] \]
A new measure of classification

How was the object classified?

\[
\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]
\]

\[
\hat{y}_O := \text{sgn} \left[ \sum_{i \in O} f(x_i) \right]
\]

Calculate value (sum) of the intersection of visualization and hand-segmented image
A new measure of classification

How was the object classified?

\[
\hat{y} = \text{sgn} \left[ \sum_{i=1}^{n} f(x_i) \right]
\]

\[
\hat{y}_O := \text{sgn} \left[ \sum_{i \in O} f(x_i) \right]
\]

Accuracy = \frac{\# \text{ correct classifications (}\hat{y}\text{)}}{\# \text{ total classifications}}

Relevant accuracy = \frac{\# \text{ correct classifications (}\hat{y}_O\text{)}}{\# \text{ total classifications}
Summary / Conclusions

Was the right classification made for the *right reason*?

- Visualization: what pixels contributed the classification, and how much they contributed.
Summary / Conclusions

Was the right classification made for the _right reason_?

- Visualization: what pixels contributed the classification, and how much they contributed.

- Constraints on classification function
- Constraints on feature extraction
- Human judgment
Summary / Conclusions

Was the right classification made for the *right reason*?

• Visualization: what pixels contributed the classification, and how much they contributed.

  - Constraints on classification function
  - Constraints on feature extraction
  - Human judgment

• Relevant accuracy: reclassify image based only on pixels from the object.

\[
\hat{y}_O = \text{sgn} \left[ \sum_{i \in O} f(x_i) + \sum_{j \in B} f(x_j) \right]
\]
Summary / Conclusions

Was the right classification made for the *right reason*?

• Visualization: what pixels contributed the classification, and how much they contributed.

- Constraints on classification function
- Constraints on feature extraction
- Human judgment

• Relevant accuracy: reclassify image based only on pixels from the object.

\[
\hat{y}_O = \text{sgn} \left[ \sum_{i \in O} f(x_i) + \sum_{j \in B} f(x_j) \right]
\]

- Same constraints as above
- Cost of hand-segmenting the object
References
