An Adaptive Interface for Active Localization

Kenji Okuma
Eric Brochu
David G. Lowe
James J. Little

The University of British Columbia
Object Localization
Problems

• How do we speed up the labelling process by selecting the most informative samples?
• Lots of data helps, but labels are expensive
• Scale of the problem
  • e.g., 3000 images:
    • over 164,000,000 windows to be evaluated for object localization
• Computation matters!
Answers
Answers

• Active learning with just Yes/No questions is not best
Answers

• Active learning with just Yes/No questions is not best

• Machine intelligence is yet effective to discover a useful portion of data
Questions

• Active learning with just Yes/No questions is not best
• Machine intelligence is yet effective to discover a useful portion of data
• An interactive system with a little amount of human effort is even better
Answers

- Active learning with just Yes/No questions is not best
- Machine intelligence is yet effective to discover a useful portion of data
- An interactive system with a little amount of human effort is even better
- Our interactive system is particularly effective when there isn’t much labelled data
pool-based active learning

Active learning systems attempt to overcome the labeling bottleneck by asking queries in the form of unlabeled instances to be labeled by an oracle: e.g., a human annotator. In this way, the active learner aims to achieve high accuracy using as few labeled instances as possible, thereby minimizing the cost of obtaining labeled data. Active learning is well-motivated in many modern machine learning problems where data may be abundant but labels are scarce or expensive to obtain.

Note that this kind of active learning is related in spirit, though not to be confused, with the family of instructional techniques by the same name in the education literature: Bonwell and Eison 1991.

1.2 Active Learning Examples

There are several scenarios in which active learners may pose queries, and there are also several different query strategies that have been used to decide which instances are most informative. In this section, I present two illustrative examples in the pool-based active learning setting: in which queries are selected from a large pool of unlabeled instances, using an uncertainty sampling query strategy, which selects the instance in the pool about which the model is least certain how to label. Sections 2 and 3 describe all the active learning scenarios and query strategy frameworks in more detail.

[Settles 2009]
pool-based active learning

From the set of unlabelled data, select the candidate data for labelling.

[Settles 2009]

(1) From the set of unlabelled data, select the candidate data for labelling.
pool-based active learning

From the set of unlabelled data, select the candidate data for labelling

Get the label from oracle

Active learning systems attempt to overcome the labeling bottleneck by asking queries in the form of unlabeled instances to be labeled by an oracle: e.g., a human annotator. In this way, the active learner aims to achieve high accuracy using as few labeled instances as possible, thereby minimizing the cost of obtaining labeled data. Active learning is well-motivated in many modern machine learning problems where data may be abundant but labels are scarce or expensive to obtain.

Note that this kind of active learning is related in spirit, though not to be confused, with the family of instructional techniques by the same name in the education literature: Bonwell and Eison, 1991.

1.2 Active Learning Examples

There are several scenarios in which active learners may pose queries, and there are also several different query strategies that have been used to decide which instances are most informative. In this section, I present two illustrative examples in the pool-based active learning setting: in which queries are selected from a large pool of unlabeled instances $U$, using an uncertainty sampling query strategy: which selects the instance in the pool about which the model is least certain how to label. Sections 2 and 3 describe all the active learning scenarios and query strategy frameworks in more detail.

(Settles 2009)
pool-based active learning

(1) From the set of unlabelled data, select the candidate data for labelling

(2) Get the label from oracle

(3) Add the labelled data to the labelled training set

Note that this kind of active learning is related in spirit, though not to be confused, with the family of instructional techniques by the same name in the education literature: Bonwell and Eison 1991.

1.2 Active Learning Examples

There are several scenarios in which active learners may pose queries, and there are also several different query strategies that have been used to decide which instances are most informative. In this section, I present two illustrative examples in the pool-based active learning setting: in which queries are selected from a large pool of unlabeled instances $U$, using an uncertainty sampling query strategy: which selects the instance in the pool about which the model is least certain how to label. Sections 2 and 3 describe all the active learning scenarios and query strategy frameworks in more detail.
pool-based active learning

(1) From the set of unlabelled data, select the candidate data for labelling
(2) Get the label from oracle
(3) Add the labelled data to the labelled training set
(4) Update classifier

[Settles 2009]
pool-based active learning

(1) From the set of unlabelled data, select the candidate data for labelling

(2) Get the label from oracle

(3) Add the labelled data to the labelled training set

(4) Update classifier

[Settles 2009]
Select candidate queries from unlabelled data
Select candidate queries from unlabelled data

1. Train state-of-the-art object detectors on PASCAL VOC 2007 data
   • 20 visual object classes
   • 9,963 images (50% train, 50% test)
Select candidate queries from unlabelled data

1. Train state-of-the-art object detectors on PASCAL VOC 2007 data
   - 20 visual object classes
   - 9,963 images (50% train, 50% test)

2. Download $3000 \times 20 = 60,000$ images from Flickr
Select candidate queries from unlabelled data

1. Train state-of-the-art object detectors on PASCAL VOC 2007 data
   - 20 visual object classes
   - 9,963 images (50% train, 50% test)
2. Download 3000 x 20 = 60,000 images from Flickr
3. Run our baseline detectors on those images and save detection windows
Select candidate queries from unlabelled data

1. Train state-of-the-art object detectors on PASCAL VOC 2007 data
   - 20 visual object classes
   - 9,963 images (50% train, 50% test)
2. Download 3000 x 20 = 60,000 images from Flickr
3. Run our baseline detectors on those images and save detection windows
4. Use active learning to rank each image based on uncertainty (select the most uncertain 10%)
Latent SVM

\[ f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \]

[Felzenszwalb et al., 2009]
Latent SVM

\[ f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \]

(Felzenszwalb et al., 2009)

- \( f_\beta \): LSVM is a classifier that scores a sample \( x \)
Latent SVM

\[ f_\beta(x) = \max_{z \in \mathcal{Z}(x)} \beta \cdot \Phi(x, z) \]

[Felzenszwalb et al., 2009]

- \( f_\beta \): LSVM is a classifier that scores a sample \( x \)
- \( \beta \): a vector of model parameters
Latent SVM

\[ f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \]

[Felzenszwalb et al., 2009]

- \( f_\beta \): LSVM is a classifier that scores a sample \( x \)
- \( \beta \): a vector of model parameters
- \( Z \): possible latent values for a sample \( x \)
Latent SVM

\[ f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \]

[Felzenszwalb et al., 2009]

- \( f_\beta \): LSVM is a classifier that scores a sample \( x \)
- \( \beta \): a vector of model parameters
- \( Z \): possible latent values for a sample \( x \)
- \( \Phi(x, z) \): feature vector
Latent SVM

[Felzenszwalb et al., 2010]
Active Learning with LSVM

- Decision function:
  \[ h(x) = f_\beta(x) + b \]

- Active learning criterion:
  \[ \text{argmin}_{x \in C} \left| \frac{h(x_i)}{\|\beta\|} \right| \]

uncertainty sampling [Lewis and Gale 1994]
Active Learning with LSVM

• Decision function:

\[ h(x) = f_\beta(x) + b \]

• Active learning criterion:

\[ \arg\min_{x \in \mathcal{C}} \left| \frac{h(x_i)}{\|\beta\|} \right| \]

uncertainty sampling [Lewis and Gale 1994]
Random v.s. Active sampling

aeroplane

bicycle

bottle

bus

chair

plant
Hardest and easiest

bicycle (.537)

Hard 5:

Easy 5:

chair (.163)

Hard 5:

Easy 5:
Active learning systems attempt to overcome the labeling bottleneck by asking queries in the form of unlabeled instances to be labeled by an oracle: e.g., a human annotator. In this way, the active learner aims to achieve high accuracy using as few labeled instances as possible, thereby minimizing the cost of obtaining labeled data. Active learning is well-motivated in many modern machine learning problems where data may be abundant but labels are scarce or expensive to obtain.

Note that this kind of active learning is related in spirit, though not to be confused, with the family of instructional techniques by the same name in the education literature: Bonwell and Eison (1991).

### 1.2 Active Learning Examples

There are several scenarios in which active learners may pose queries, and there are also several different query strategies that have been used to decide which instances are most informative. In this section, I present two illustrative examples in the pool-based active learning setting: in which queries are selected from a large pool of unlabeled instances, using an uncertainty sampling query strategy: which selects the instance in the pool about which the model is least certain how to label.

Sections 2 and 3 describe all the active learning scenarios and query strategy frameworks in more detail.
Human annotation: Yes / No
Human annotation: Maybe / No Target
Demo
System

• Python / C
• Utilize multi-core CPUs to speed up the interaction process
• Image features are precomputed for real time object detection
• real-time interaction for updating and determining the next query windows upon receiving new labels
pool-based active learning

(1) From the set of unlabelled data, select the candidate data for labelling

(2) Get the label from oracle

(3) Add the labelled data to the labelled training set

(4) Update classifier

[Settles 2009]
Active learning systems attempt to overcome the labeling bottleneck by asking queries in the form of unlabeled instances to be labeled by an oracle: e.g., a human annotator. In this way, the active learner aims to achieve high accuracy using as few labeled instances as possible, thereby minimizing the cost of obtaining labeled data. Active learning is well-motivated in many modern machine learning problems where data may be abundant but labels are scarce or expensive to obtain.

Note that this kind of active learning is related in spirit, though not to be confused, with the family of instructional techniques by the same name in the education literature: Bonwell and Eison 1991.

1.2 Active Learning Examples

There are several scenarios in which active learners may pose queries, and there are also several different query strategies that have been used to decide which instances are most informative. In this section, I present two illustrative examples in the pool-based active learning setting: in which queries are selected from a large pool of unlabeled instances, using an uncertainty sampling query strategy, which selects the instance in the pool about which the model is least certain how to label. Sections 2 and 3 describe all the active learning scenarios and query strategy frameworks in more detail.

[Settles 2009]
Experiments

- PASCAL VOC 2007 test data (4,981 images)
- Three approaches:
  - **LSVM+HOG**: baseline detectors trained on PASCAL VOC 2007
  - **ALORquery**: baseline + active learning with LSVM (common uncertainty sampling criteria)
  - **ALORfull**: baseline + active learning with LSVM + human interactions
## Results on 20 visual categories

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. user time (minutes)</th>
<th>Avg. # of additional Labels</th>
<th>Mean AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pos</td>
<td>neg</td>
</tr>
<tr>
<td>LSVM + HOG</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ALORquery</td>
<td>20</td>
<td>140</td>
<td>58</td>
</tr>
<tr>
<td>ALORfull</td>
<td>40</td>
<td>284</td>
<td>72</td>
</tr>
</tbody>
</table>

March 5, VISAPP 2011
Category-wise Breakdown of Gain in Average Precision

VOC2007: Category-wise Breakdown of Gain in Average Precision

Gain in Average Precision per category (%) (Baseline+ALOR - Baseline)

Categories: bus, sheep, bird, cow, cat, table, dog, sofa, plant, bottle, train, mbike, car, bicycle, chair, aerop., tv, boat, person, horse.

Heterogeneous shapes and sizes

boat

horse
**ALORfull**: Adaptive Interface (query by machine + human)  
V.S.  
**ALORquery**: Active Learning with LSVM (query by machine only)
**ALORfull**: Adaptive Interface (query by machine + human)  
**V.S.**  
**ALORquery**: Active Learning with LSVM (query by machine only)

Figure 5. Performance comparison of ALORfull and ALORquery: Each point represents the average number of additional positive labels and the average training time per category. By allowing a user to add additional queries, the performance is consistently better. Note that just 53 labels from ALORfull outperform 140 labels from ALORquery.

Figure 6 presents the gain of our best result for each category in the average precision. From our experiments, we can make several observations. First, in most of the categories, our active learning interface allows a user to quickly improve the performance of the baseline detectors. Second, our user interface also allows a user to achieve a better performance than would be obtained by a simpler learning approach in which a user answers Yes/No/Maybe queries for selected windows. A lot of difficult machine-selected queries that a user is not sure about, for example, can be easily corrected with our interface. Third, with less than 45 minutes of user input, we can achieve significant performance improvement even over the best competition results. The PASCAL competition has a section in which users can provide their own data, but the difficulty of collecting such data means there have seldom been entries in that section. Our active learning approach and GUI would enable users to efficiently collect useful data for improved performance in such competitions or for real-world applications.

In the boat, horse, and person categories, our active learning approach did not provide much improvement for the relatively small additional amount of training data. We observe that both the boat and horse categories have a particularly heterogeneous dataset in both the size and shape of the object. Some of those examples are presented in Figure 7. The person category differs in that it already has so many object labels (4606) as opposed to the median (346) that it likely requires many more labels than the few hundred we provide to improve performance.

6 CONCLUSION

We have presented an active learning system for object recognition and localization. This work differs from active learning work for image classification in that instead of learning a single classification for the image, our model can identify and localize objects, including multiple instances of the object of interest in a single image. Our experiments demonstrate that the active learning approach reduces the number of labels required to train an object classifier without reducing performance over state-of-the-art classification schemes. It also greatly reduces the human effort required to select image regions containing the object of interest by automatically finding the most useful windows in an image. Our system is fast enough to be used interactively, and we demonstrate a prototype GUI for active learning of object locations, which uses image windows to guide human labelling. While our experiments show that our actively trained latent SVM with HOG descriptors works well with active learning, the system does not depend on a specific classifier or feature set. If other classifiers, such as AdaBoost, or features are found to be more suitable to a domain, we can incorporate them into our framework. We also believe that other aspects of object classification can benefit from active learning as the expense of labelling is a ubiquitous problem in machine learning. Fast, efficient labelling can mean cheaper experiments, faster development time, and higher performance flexible object detectors.

7 ACKNOWLEDGMENT

The research has been supported by the Natural Sciences and Engineering Research Council of Canada and the GEOIDE Network of Centres of Excellence, Terrapoint Canada’s division of Ambercore.
ALORfull: Adaptive Interface (query by machine + human)  
V.S.  
ALORquery: Active Learning with LSVM (query by machine only)

Figure 5. Performance comparison of ALORfull and ALORquery: Each point represents the average number of additional positive labels and the average training time per category. By allowing a user to add additional queries, the performance is consistently better. Note that just 53 labels from ALORfull outperform 140 labels from ALORquery.

Fig. 6 presents the gain of our best result for each category in the average precision. From our experiments, we can make several observations. First, in most of the categories, our active learning interface allows a user to quickly improve the performance of the baseline detectors. Second, our user interface also allows a user to achieve a better performance than would be obtained by a simpler learning approach in which a user answers Yes/No/Maybe queries for selected windows. A lot of difficult machine-selected queries that a user is not sure about, for example, can be easily corrected with our interface. Third, with less than 46 minutes of user input, we can achieve significant performance improvement even over the best competition results. The PASCAL competition has a section in which users can provide their own data, but the difficulty of collecting such data means there have seldom been entries in that section. Our active learning approach and GUI would enable users to efficiently collect useful data for improved performance in such competitions or for real-world applications.

6 CONCLUSION

We have presented an active learning system for object recognition and localization. This work differs from active learning work for image classification in that instead of learning a single classification for the image, our model can identify and localize objects, including multiple instances of the object of interest in a single image. Our experiments demonstrate that the active learning approach reduces the number of labels required to train an object classifier without reducing performance over state-of-the-art classification schemes. It also greatly reduces the human effort required to select image regions containing the object of interest by automatically finding the most useful windows in an image. Our system is fast enough to be used interactively, and we demonstrate a prototype GUI for active learning of object locations, which uses image windows to guide human labelling. While our experiments show that our actively trained latent SVM with HOG descriptors works well with active learning, the system does not depend on a specific classifier or feature set. If other classifiers, such as AdaBoost, or features are found to be more suitable to a domain, we can incorporate them into our framework. We also believe that other aspects of object classification can benefit from active learning, as the expense of labelling is a ubiquitous problem in machine learning. Fast, efficient labelling can mean cheaper experiments, faster development time, and higher-performance flexible object detectors.

7 ACKNOWLEDGMENT

The research has been supported by the Natural Sciences and Engineering Research Council of Canada and the GEOIDE Network of Centres of Excellence, Terrapoint Canada (a division of Ambercore.)

a user can choose more useful labels with a little effort!
## Query Analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. Training Time (mins)</th>
<th>user selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>machine selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>Avg. additional labels pos</th>
<th>Avg. additional labels neg</th>
<th>Mean AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALORquery 300 images</td>
<td>20</td>
<td>0</td>
<td>NA</td>
<td>100</td>
<td>0.283</td>
<td>140.1</td>
<td>58.25</td>
<td>28.3</td>
</tr>
<tr>
<td>ALORfull 100 images</td>
<td>13</td>
<td>42</td>
<td>0.643</td>
<td>58</td>
<td>0.200</td>
<td>106.85</td>
<td>20.15</td>
<td>29.4</td>
</tr>
<tr>
<td>ALORfull 300 images</td>
<td>40</td>
<td>43</td>
<td>0.690</td>
<td>57</td>
<td>0.296</td>
<td>284.8</td>
<td>72.1</td>
<td>30.6</td>
</tr>
</tbody>
</table>
Query Analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. Training Time (mins)</th>
<th>user selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>machine selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>Avg. additional labels pos</th>
<th>Avg. additional labels neg</th>
<th>Mean AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALORquery 300 images</td>
<td>20</td>
<td>0</td>
<td>NA</td>
<td>100</td>
<td>0.283</td>
<td>140.1</td>
<td>58.25</td>
<td>28.3</td>
</tr>
<tr>
<td>ALORfull 100 images</td>
<td>13</td>
<td>42</td>
<td>0.643</td>
<td>58</td>
<td>0.200</td>
<td>106.85</td>
<td>20.15</td>
<td>29.4</td>
</tr>
<tr>
<td>ALORfull 300 images</td>
<td>40</td>
<td>43</td>
<td>0.690</td>
<td>57</td>
<td>0.296</td>
<td>284.8</td>
<td>72.1</td>
<td>30.6</td>
</tr>
</tbody>
</table>

- Humans tend to pick queries that are further away from the decision boundary of LSVM (mostly missed detections!)
Query Analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. Training Time (mins)</th>
<th>user selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>machine selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>Mean %Ps</th>
<th>Avg. additional labels pos</th>
<th>Avg. additional labels neg</th>
<th>Mean AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALORquery 300 images</td>
<td>20</td>
<td>0</td>
<td>NA</td>
<td>100</td>
<td>0.283</td>
<td>140.1</td>
<td>58.25</td>
<td>28.3</td>
<td></td>
</tr>
<tr>
<td>ALORfull 100 images</td>
<td>13</td>
<td>42</td>
<td>0.643</td>
<td>58</td>
<td>0.200</td>
<td>106.85</td>
<td>20.15</td>
<td>29.4</td>
<td></td>
</tr>
<tr>
<td>ALORfull 300 images</td>
<td>40</td>
<td>43</td>
<td>0.690</td>
<td>57</td>
<td>0.296</td>
<td>284.8</td>
<td>72.1</td>
<td>30.6</td>
<td></td>
</tr>
</tbody>
</table>

- Humans tend to pick queries that are further away from the decision boundary of LSVM (mostly missed detections!)
- With humans in a loop, it requires less training data and less training time
## Query Analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. Training Time (mins)</th>
<th>user selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>machine selected queries proportion (%)</th>
<th>Avg. distance</th>
<th>Avg. additional labels</th>
<th>Mean AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALORquery 300 images</td>
<td>20</td>
<td>0</td>
<td>NA</td>
<td>100</td>
<td>0.283</td>
<td>140.1</td>
<td>58.25</td>
</tr>
<tr>
<td>ALORfull 100 images</td>
<td>13</td>
<td>42</td>
<td>0.643</td>
<td>58</td>
<td>0.200</td>
<td>106.85</td>
<td>20.15</td>
</tr>
<tr>
<td>ALORfull 300 images</td>
<td>40</td>
<td>43</td>
<td>0.690</td>
<td>57</td>
<td>0.296</td>
<td>284.8</td>
<td>72.1</td>
</tr>
</tbody>
</table>

- Humans tend to pick queries that are further away from the decision boundary of LSVM (mostly missed detections!)
- With humans in a loop, it requires less training data and less training time
Conclusion
Conclusion

• Our empirical study showed that an interactive system for active localization is quite effective
Conclusion

- Our empirical study showed that an interactive system for active localization is quite effective.
- Our system does not depend on a specific classifier or feature set.
Conclusion

• Our empirical study showed that an interactive system for active localization is quite effective

• Our system does not depend on a specific classifier or feature set

• For already high performance classifiers, useful positive labels are ones that are hard to classify as positive (not so obvious for us!!)
Conclusion

• Our empirical study showed that an interactive system for active localization is quite effective.

• Our system does not depend on a specific classifier or feature set.

• For already high performance classifiers, useful positive labels are ones that are hard to classify as positive (not so obvious for us!!)

• Fast, efficient labelling can mean cheaper experiments, faster development time, and higher-performance flexible object detectors.
Thanks