**Problem**
Image labels alone are insufficient supervision for learning complex visual recognition tasks.

**Our Idea**
- Annotators should not only assign class labels (the “what”), but also give a rationale indicating their reasoning behind the label (the “why”).
- We propose two modes for visual rationales:
  - **Spatial**: draw polygons around important image regions
  - **Attribute**: name attributes most influential in label choice

**Annotation task**: Is the skater’s form good? How can you tell?

**SVM Training with Contrast Examples**
- Require classifier to treat contrast example that lacks the important features as “less positive” than the original.
- We adopt the SVM objective developed by Zaidan et al., [HLT 2007] for sentiment analysis in documents:

\[
\begin{align*}
\text{Minimize:} & \quad \frac{1}{2} \|w\|^2 + c \left( \sum_{i=1}^{N} \xi_i \right) + c_c \left( \sum_{i=1}^{N} \eta_i \right) \\
\text{Subject to:} & \quad \forall i \left( y_i \left( w^T x_i \right) \geq 1 - \xi_i \right) \\
& \quad \forall i \left( y_i \left( w^T x_i - \left( w^T \tilde{x}_i \right) \right) \geq \mu (1 - y_i) - \xi_i \right) \\
& \quad \xi_i, \eta_i \geq 0
\end{align*}
\]

where \( x_i \) is the \( i \)-th training example, \( y_i \) is its corresponding contrast example, and \( y_i \) is the class label \{1, -1\}.  

**Results: Hot or Not?**
- Test our spatial rationales on hotornot.com using provided ratings +104 MTurk rationales
- **Task**: Classify male/female as “hot” (top 25%) or “not” (bottom 25%)

**Results: Scene Categorization**
- Test our spatial rationales on 15 Scene Categories dataset with annotations from 545 unique MTurk workers
- **Task**: Name the scene type

**Results: Public Figure Attractiveness**
- Test our attribute rationales on PubFig dataset
- **Task**: Classify public figure as attractive or not

**Conclusions**
- The “why” matters
- Positive results in multiple domains
- Rationales give deeper insight than a class label alone, especially useful in subjective tasks