**VISALOGY: Answering Visual Analogy Questions**

Fereshteh Sadeghi, Larry Zitnick, Ali Farhadi
fsadeghi@cs.uw.edu, larryz@microsoft.com, afarhad2@cs.uw.edu

---

**What is Analogy?**

(from Greek ἀναλογία, analogia, “proportion”) is a cognitive process of transferring information or meaning from a particular subject (the analogue or source) to another (the target) —wiki

**Analogy Questions**

A is to B as C is to what?

\[ A : B :: C : ? \]

**Key to Answer analogy Questions**

- Understand the semantic transformation that takes A to B
- Apply the same transformation on C to obtain the answer

**Actual Results**

Learn to Generalize

- NO textual label is given during training!
- Only analogical image quadruples are seen during training

**Formulating Visual Analogy**

Image1 to Image2 is similar to Image3 to what image?

\[ \begin{align*}
I_1 &= I^{(c_1, p_1)}; & I_2 &= I^{(c_2, p_2)}; & I_3 &= I^{(c_3, p_1)}; & I_4 &= I^{(c_4, p_2)}
\end{align*} \]

I: Image  c: Category  p: Property

- Discover mapping T from I1 to I2
- Learn parameters \( \theta \)
- Apply T on I3 and search for I4

\[ X_{12} \approx X_{34} \]

**Double margin contrastive loss to learn generalizable embedding**

\[ \min_{\theta} \max_{x_{12}, x_{34}} \left[ \left| \frac{1}{n} \sum_{i=1}^{n} \left( m_{N} \mathbb{1}_{x_{12, i} < x_{34, i}} + m_{P} \mathbb{1}_{x_{12, i} > x_{34, i}} \right) \right| \right] \]

**At test time:**

\[ \text{rank}_{i} = \frac{T(I_{12, i}) \cdot T(I_{34, i})}{\left( T(I_{12, i}) \right)^{\frac{1}{2}} \left( T(I_{34, i}) \right)^{\frac{1}{2}}} \]

---

**Experiments And Results**

**Qualitative**

- Analogy questions using natural images:
  - Various Attributes & Actions (122 phrases (14 categories, 22 properties)

**Quantitative**

- Analogy questions using synthesized images of chairs
  - Styles as categories, Poses as properties (~1400 styles, 31 poses)