VISALOGY

Visual Analogy Question Answering

Fereshteh Sadeghi, Larry Zitnick, Ali Farhadi

Analogy

...(from Greek ἀναλογία, analogía, "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

Walk is to Legs as

A. Gleam is to Eyes

B. Chew is to Mouth

C. Dress is to Hem

D. Cover is to Book

E. Grind is to Nose

Analogy questions

Walk is to Legs as

A. Gleam is to Eyes

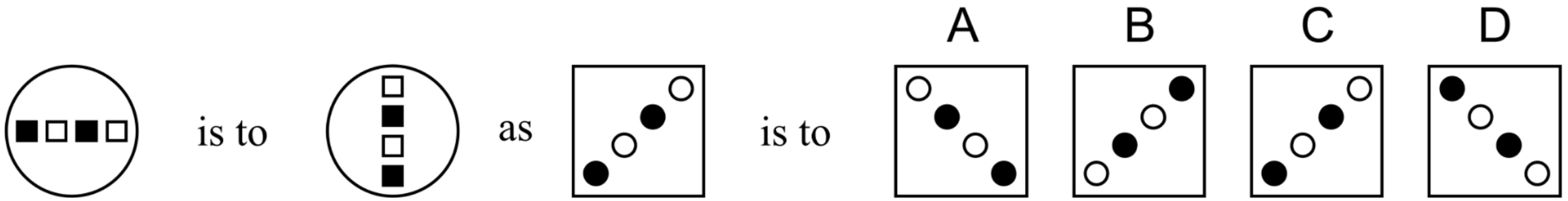
B. Chew is to Mouth

C. Dress is to Hem

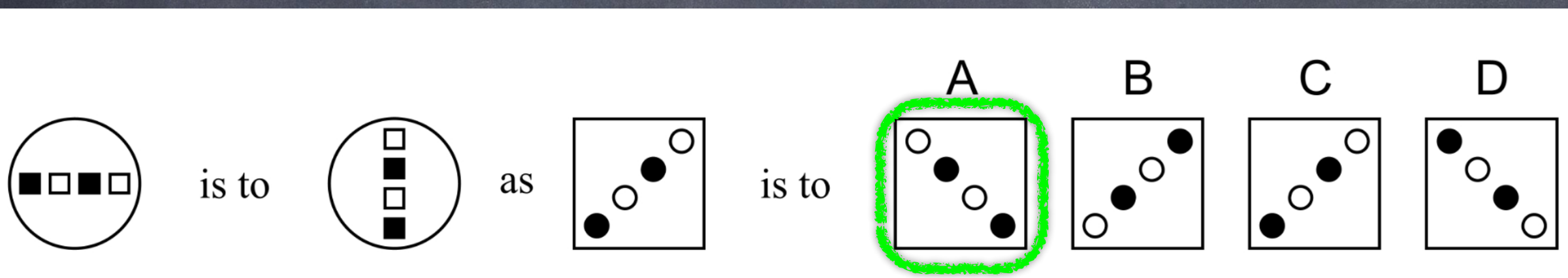
D. Cover is to Book

E. Grind is to Nose

Analogy questions



Analogy questions



A : B :: C : ?

A : B :: C : ?

A is to B as C is to ?

Brown Bear : White Bear : : Brown Dog : ?

: :

: :

: : ?

::  ::  :: ?

Possible Answers








Discover the Transformation

:

::

:

?

Discover the Transformation

:

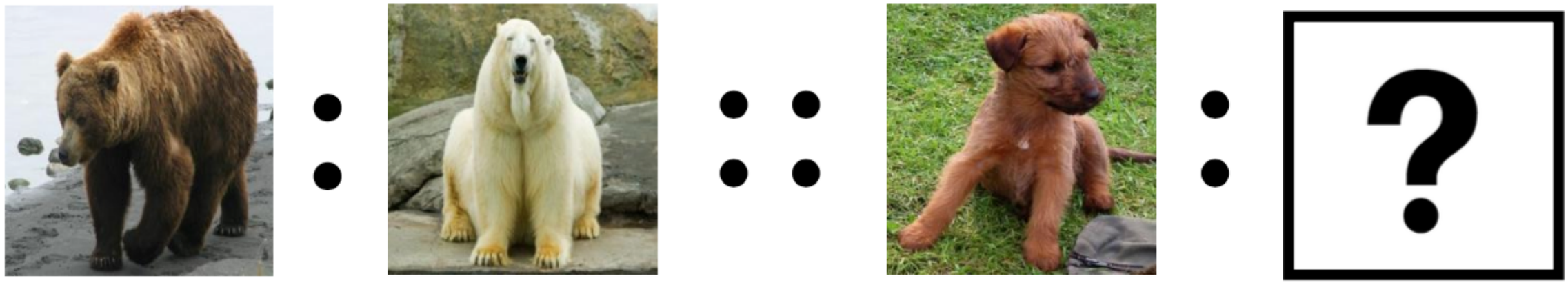
::

:

?

Apply the same Transformation

Discover the Transformation

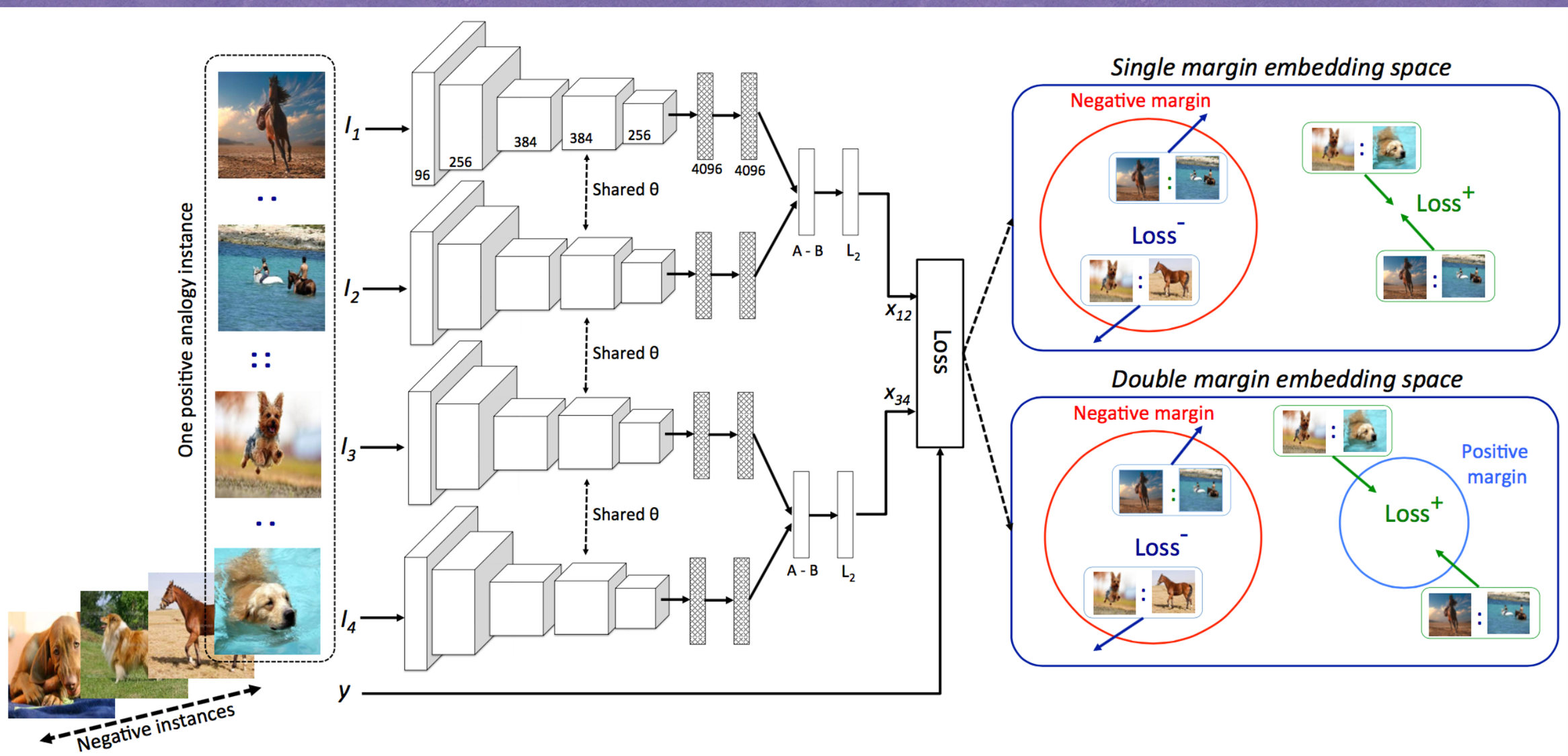


Apply the same Transformation

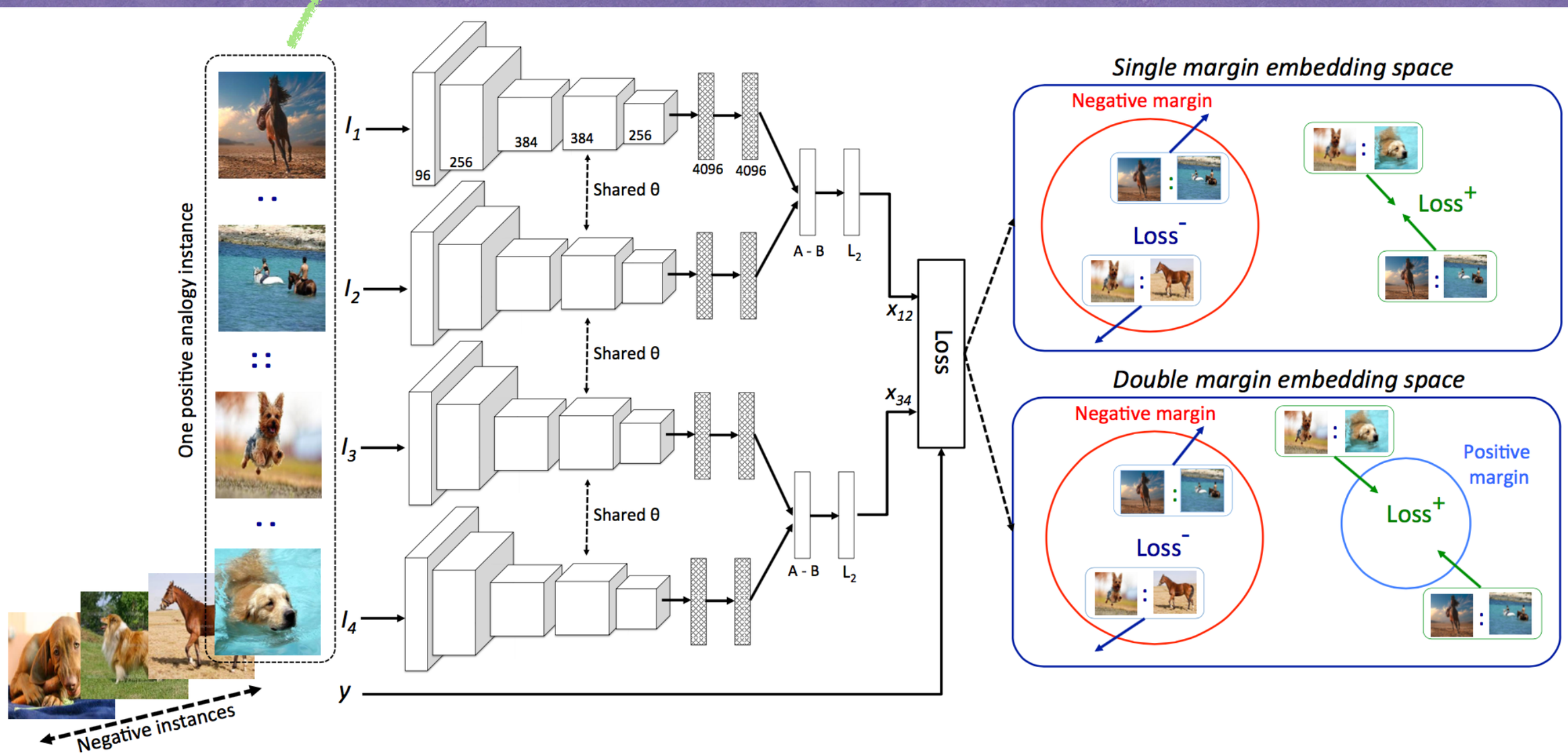
Learn an embedding:

- Image pairs with similar transformations be close in the embedding space

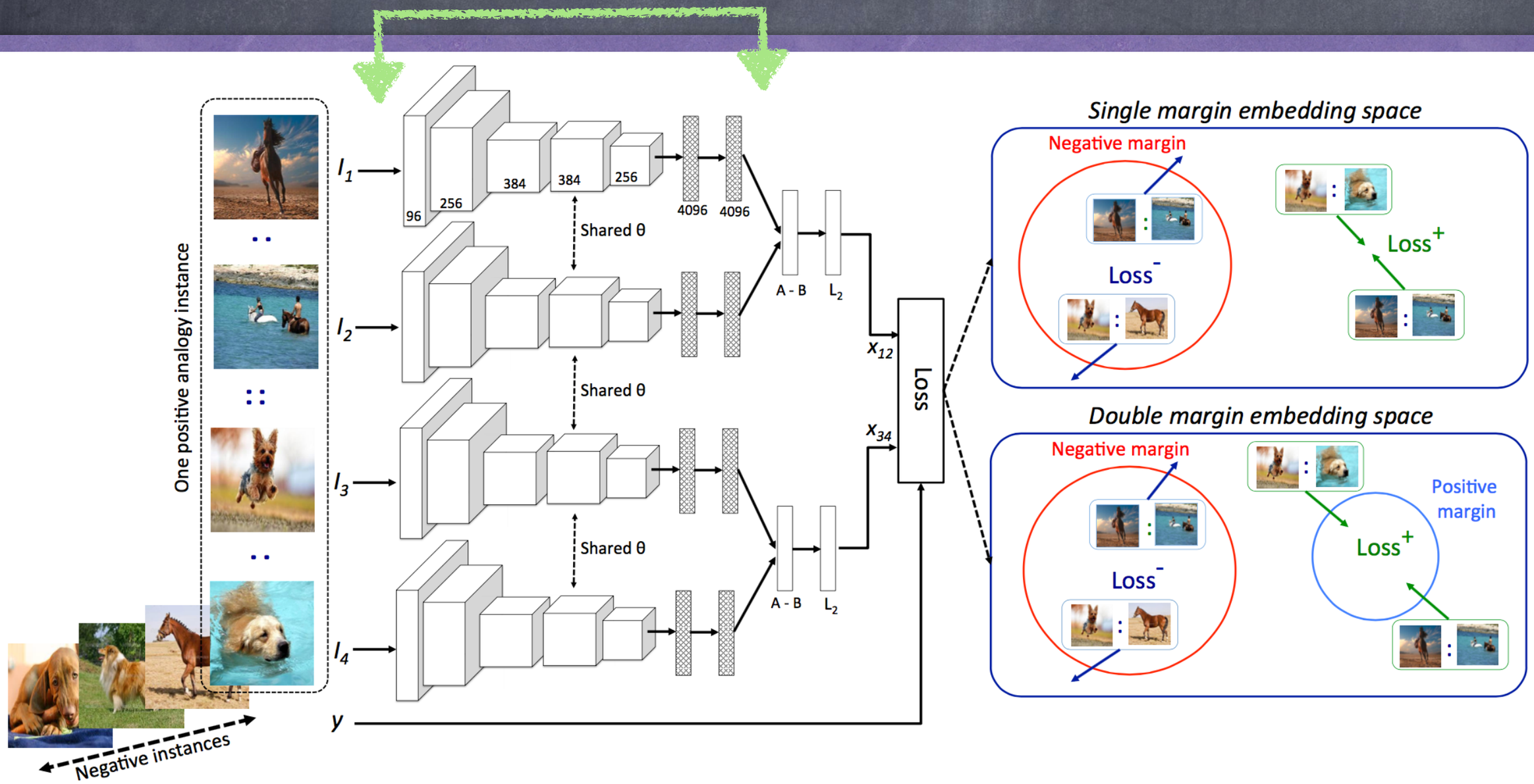
Siamese Quadruple Network



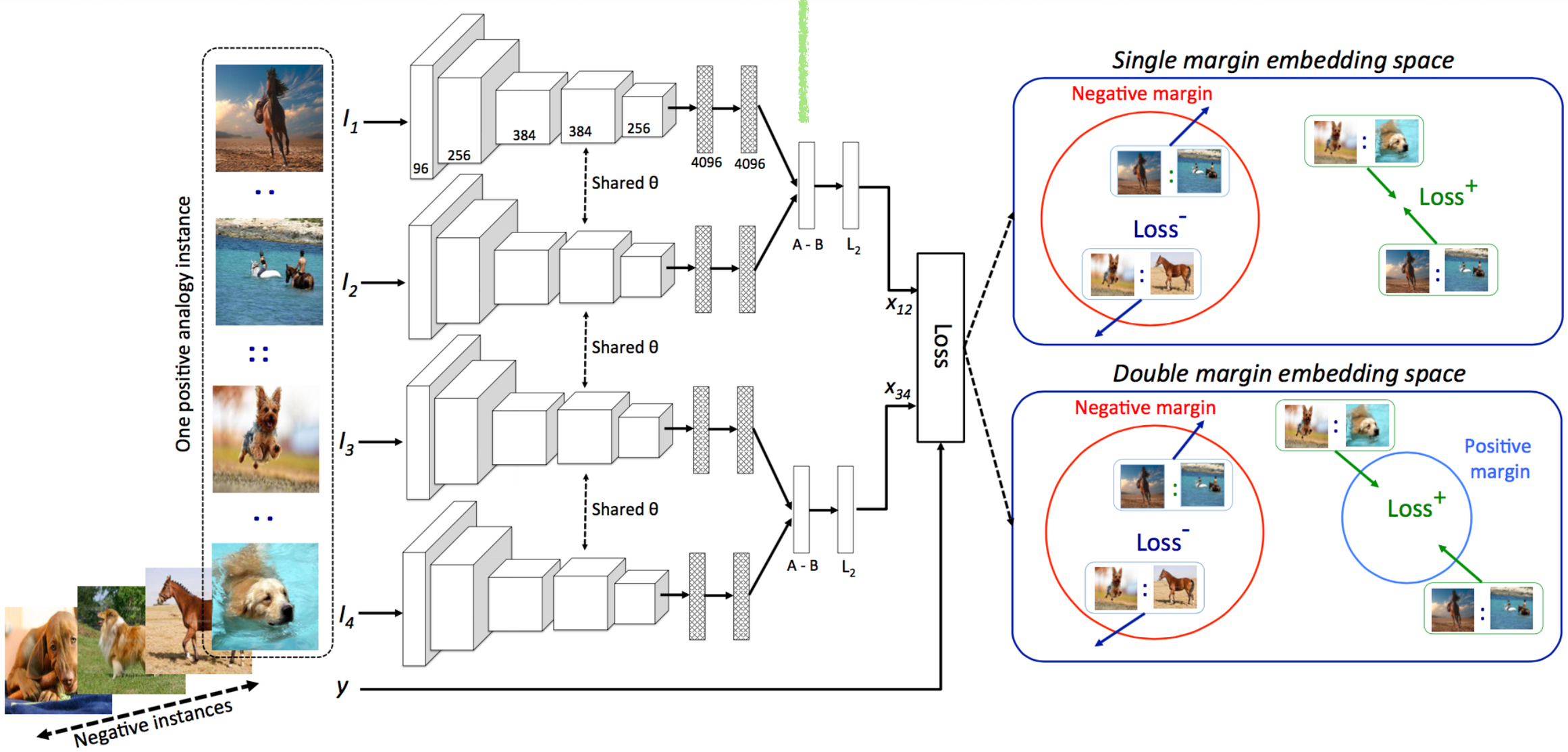
Input: $[I_1^{(c_i,p_1)} : I_2^{(c_i,p_2)} :: I_3^{(c_o,p_1)} : I_4^{(c_o,p_2)}]$



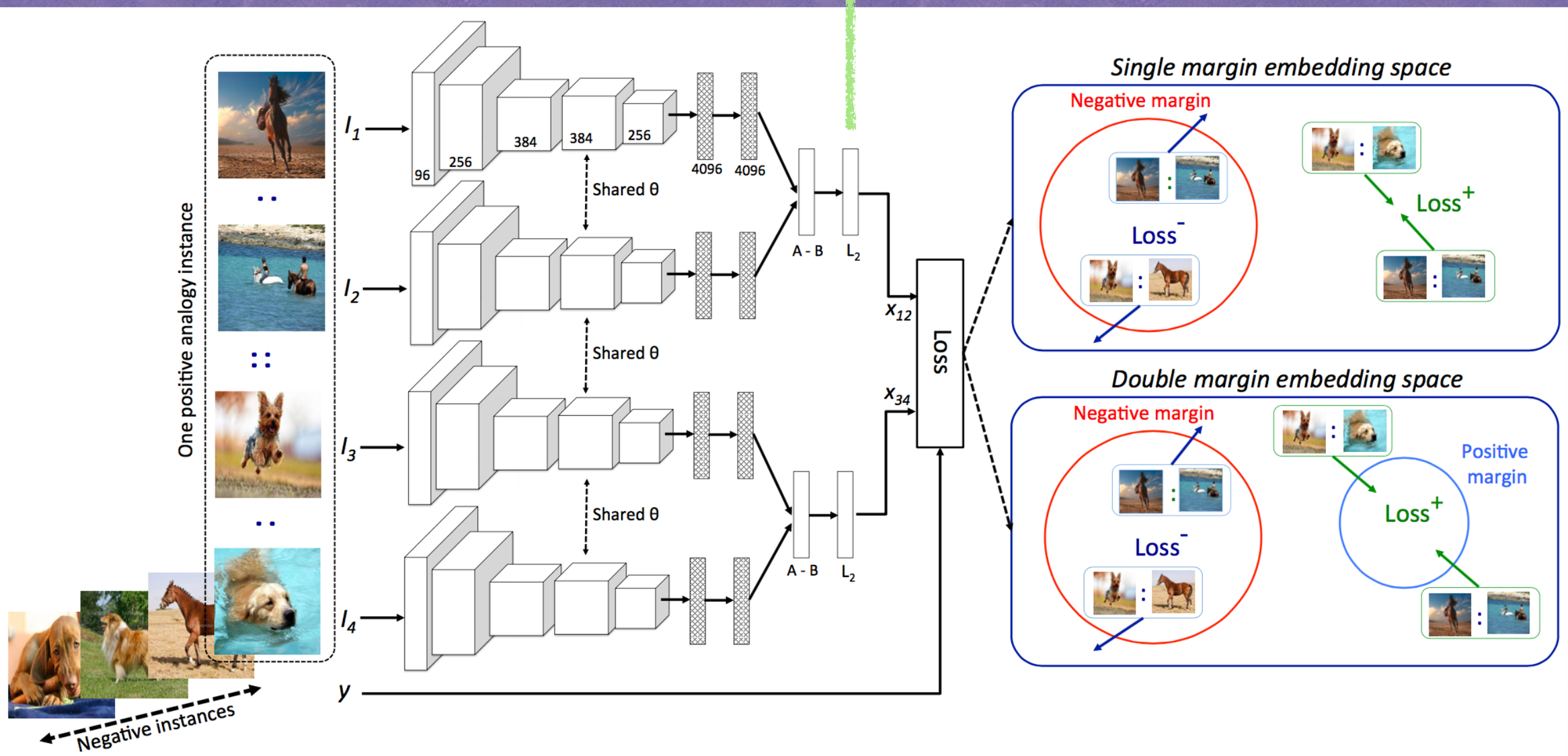
AlexNet with shared parameters



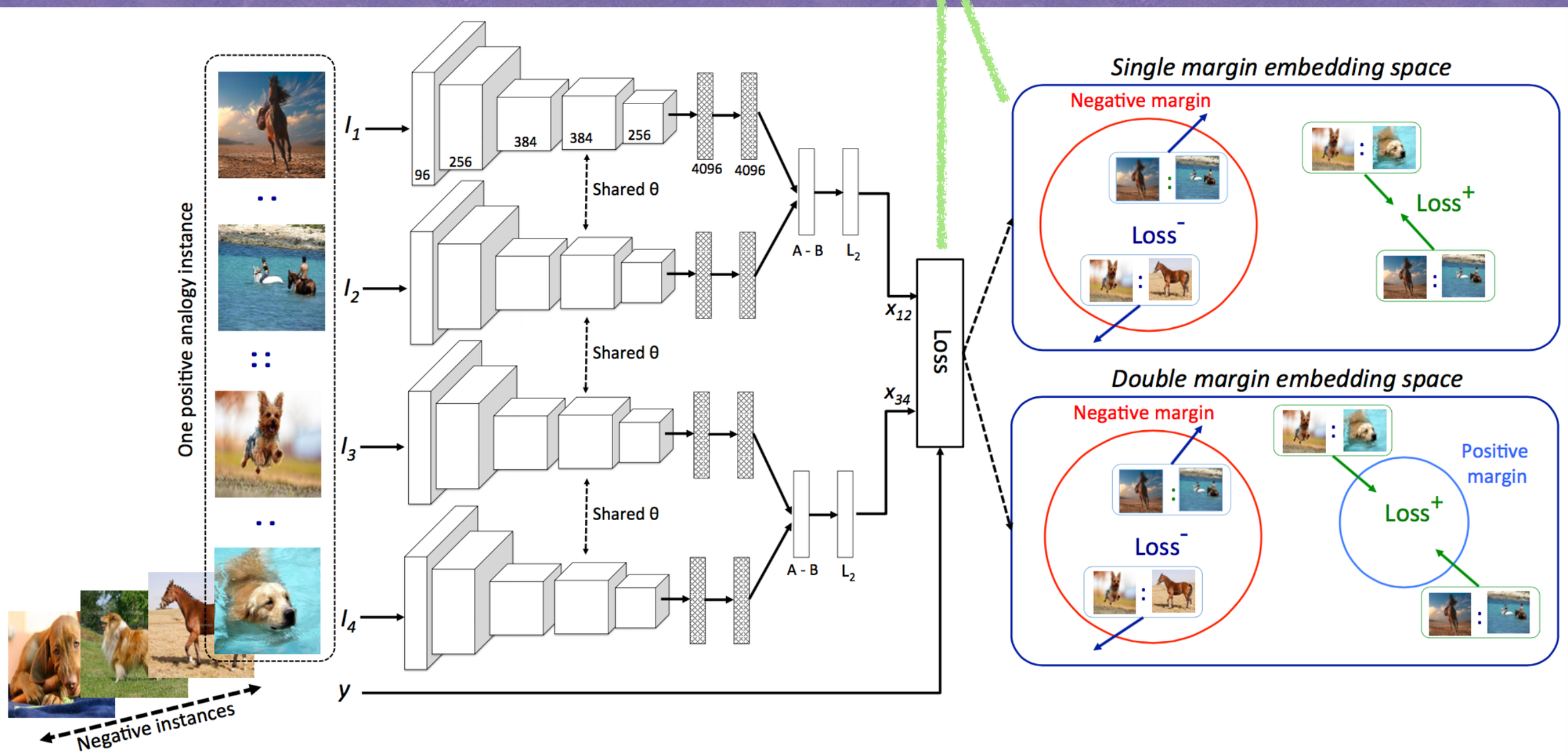
Subtraction Layer: A - B



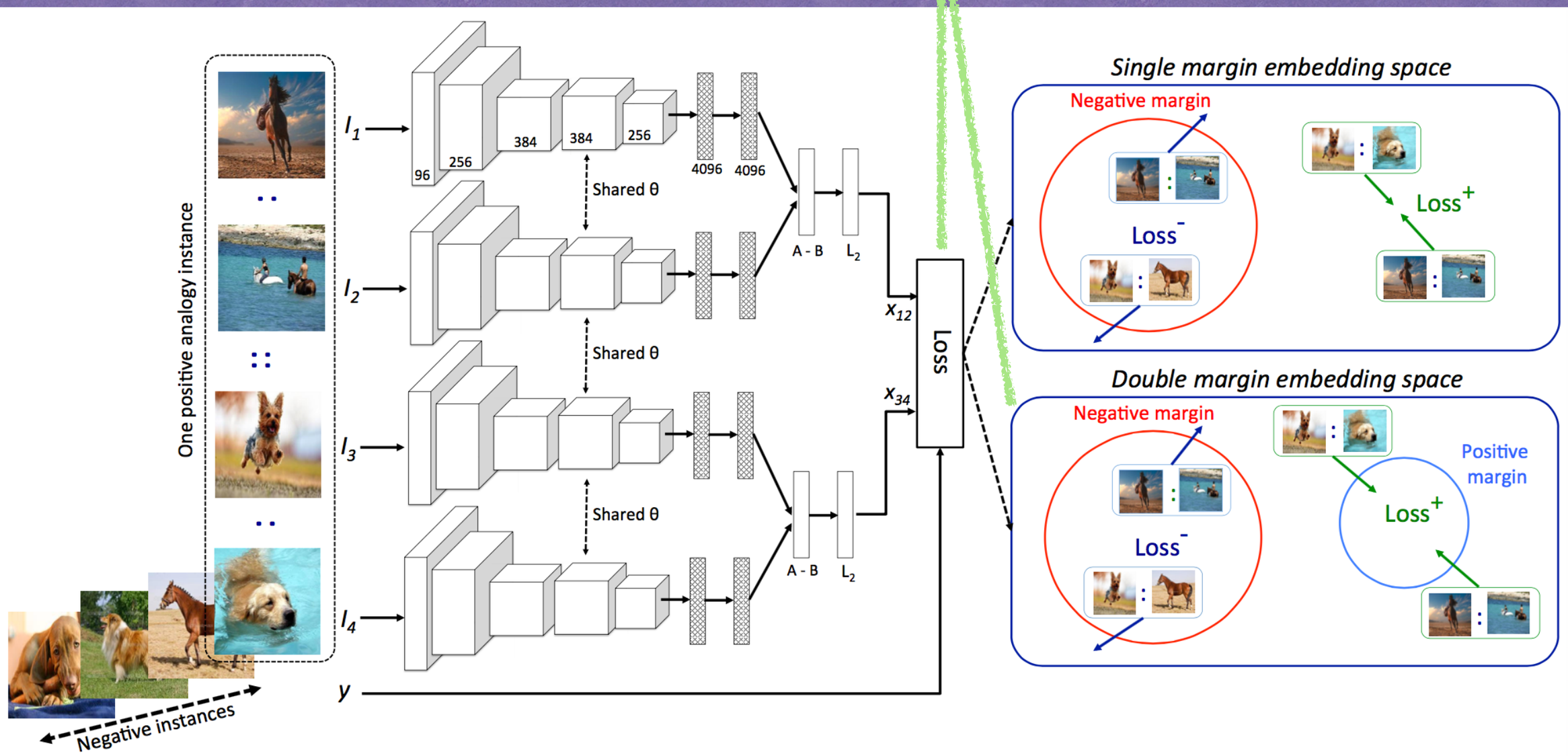
L2 Normalization Layer



Loss: $\mathcal{L}^m(x_{12}, x_{34}) = y||x_{12} - x_{34}|| + (1-y)\max(m - ||x_{12} - x_{34}||, 0)$

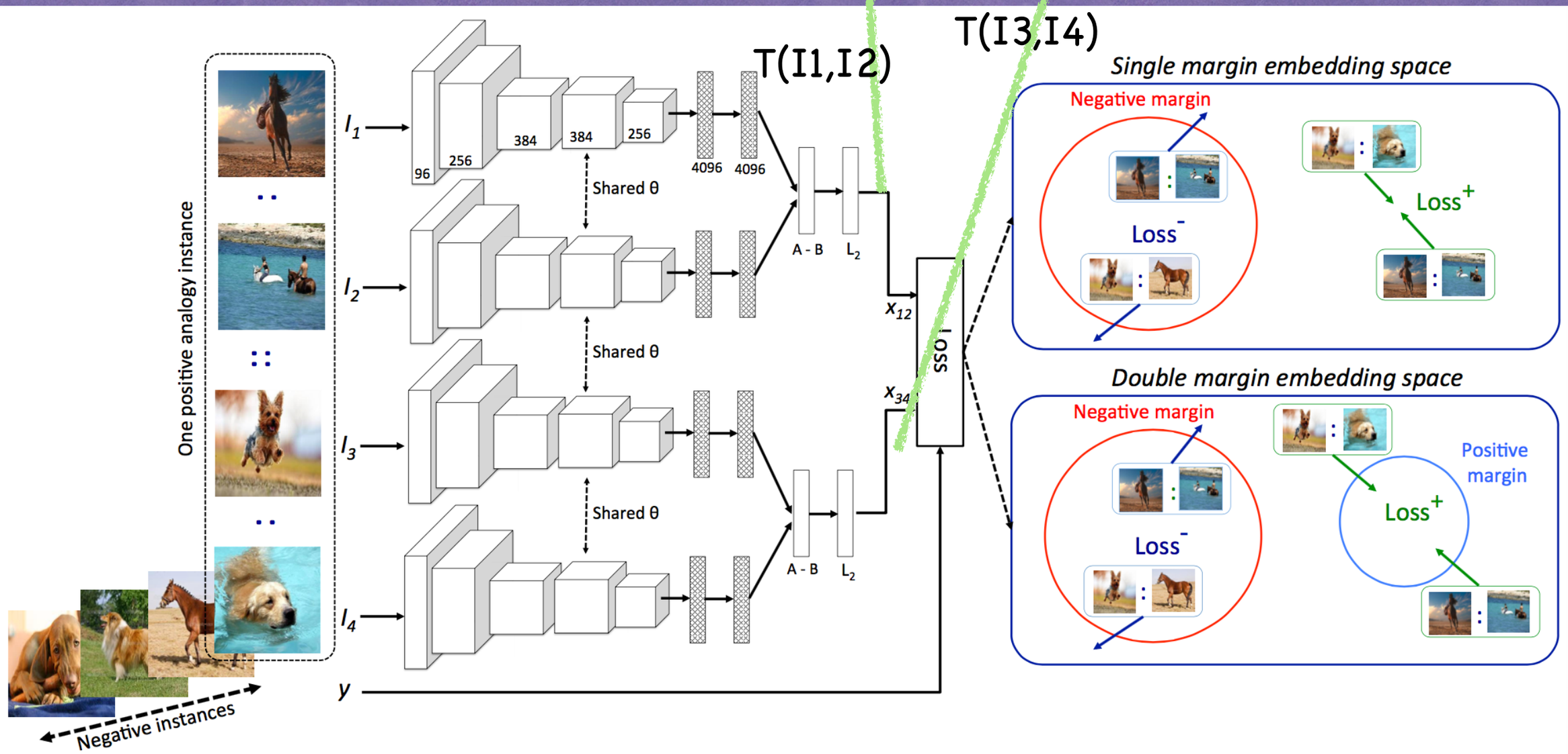


Loss: $\mathcal{L}^{m_P,m_N}(x_{12},x_{34}) = y\max(||x_{12}-x_{34}||-m_P,0)+(1-y)\max(m_N-||x_{12}-x_{34}||,0)$



Test by retrieval:

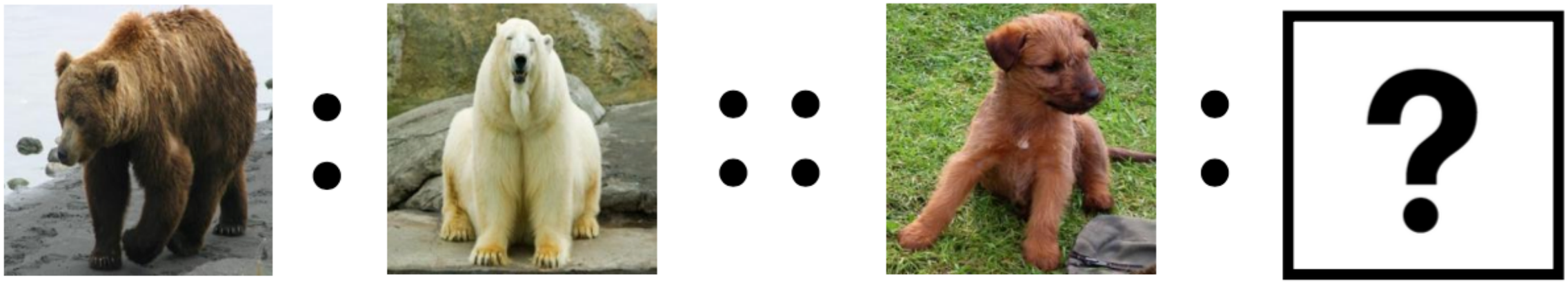
$$rank_i = \frac{T(I_1, I_2).T(I_3, I_i)}{||T(I_1, I_2)||.||T(I_3, I_i)||}, \quad i \in 1, ..., n$$



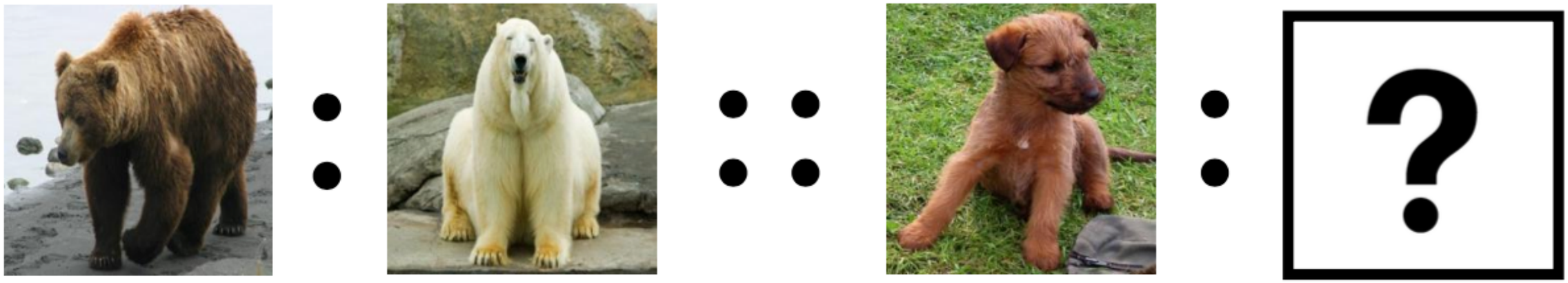

: :

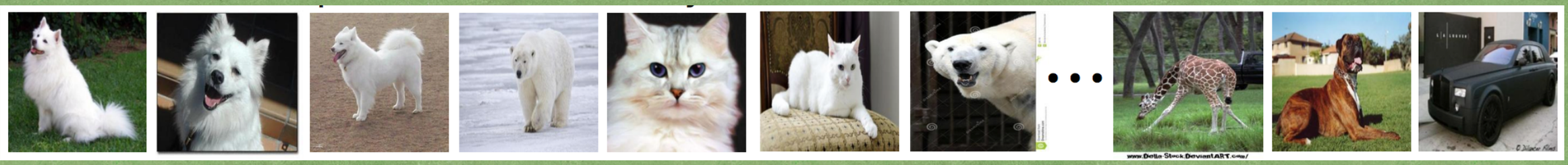
: :

: ?

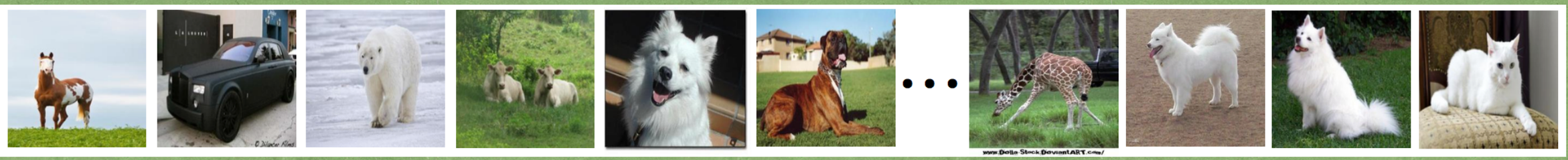


ranked selections by quadruple network





ranked selections by quadruple network



ranked selections by AlexNet

Testing…

- Natural images

  A. Various Attributes

  B. Various Actions

  - 112 phrases (14 categories, 22 properties)

Testing…

- Natural images
  - A. Various Attributes
  - B. Various Actions
    - 112 phrases (14 categories, 22 properties)

- Synthetic chairs
  - A. Various Styles
  - B. Various pose
    - ~1400 styles , 31 poses

Testing...

- Scenarios

  A. Seen Analogies

  - Test questions with formats that are seen during training

  B. Unseen Analogies

  - Test questions with formats that are unseen during training

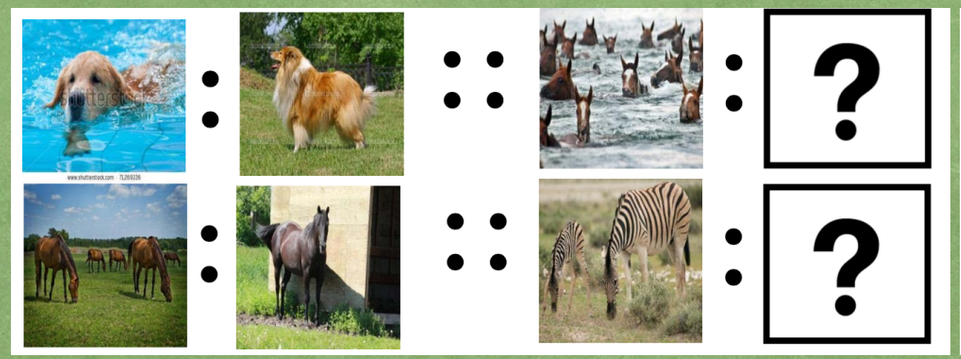
Qualitative Results (Natural images)

Questions

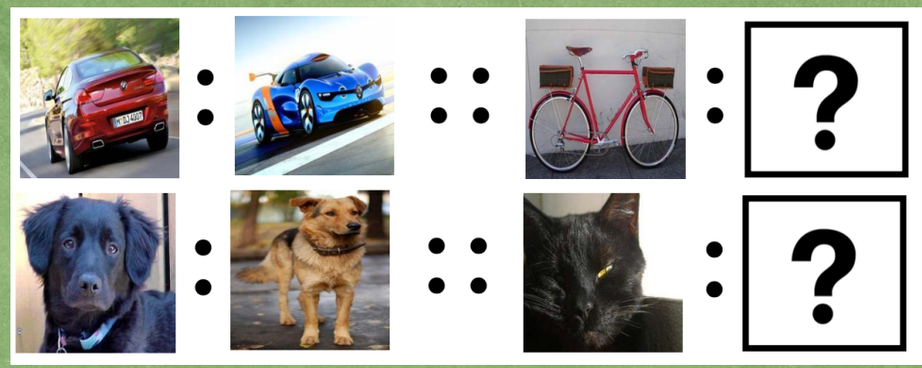
Ours

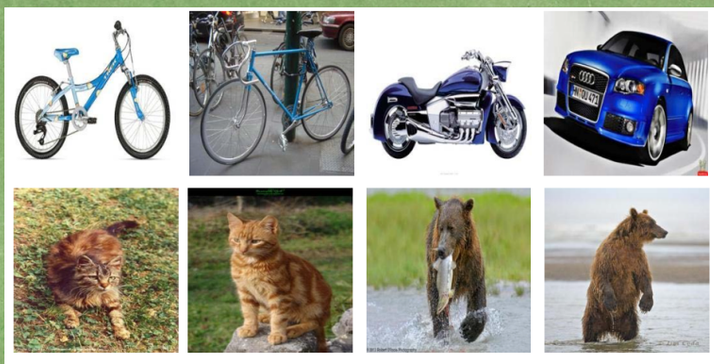
Baseline

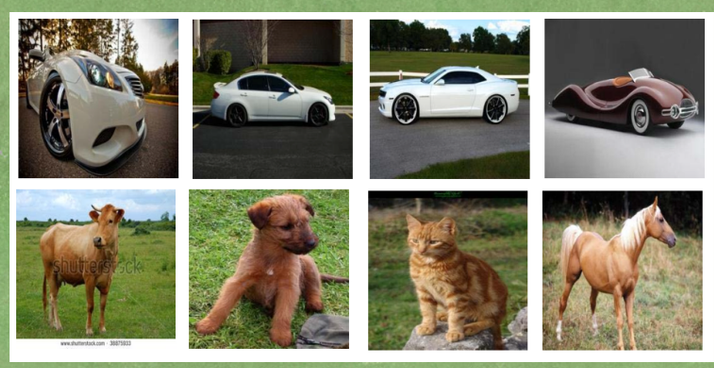
Action



Attribute

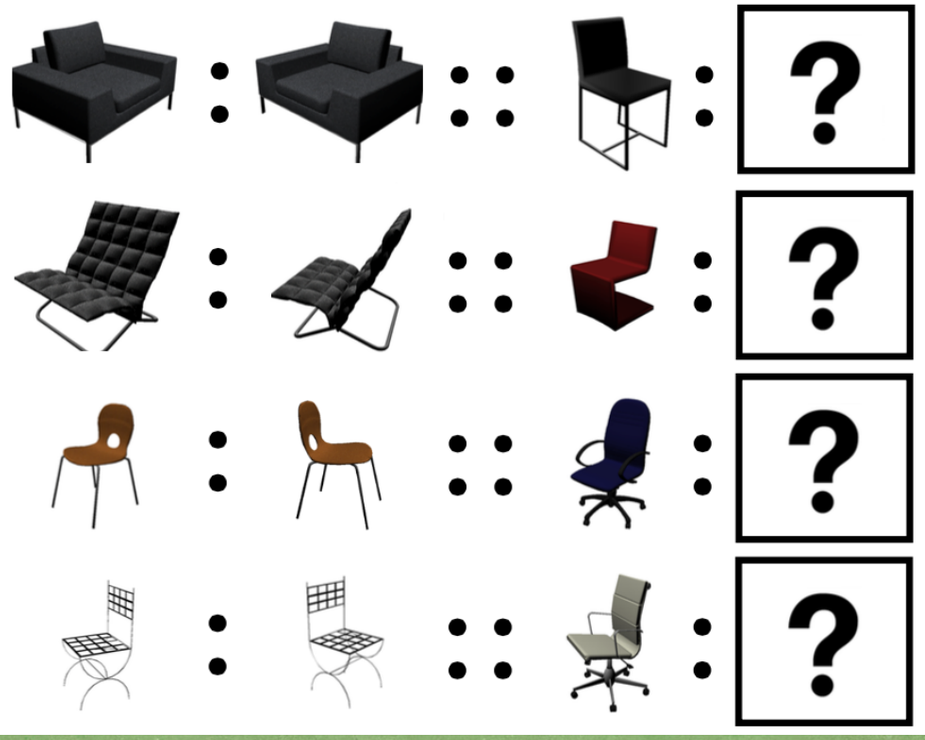






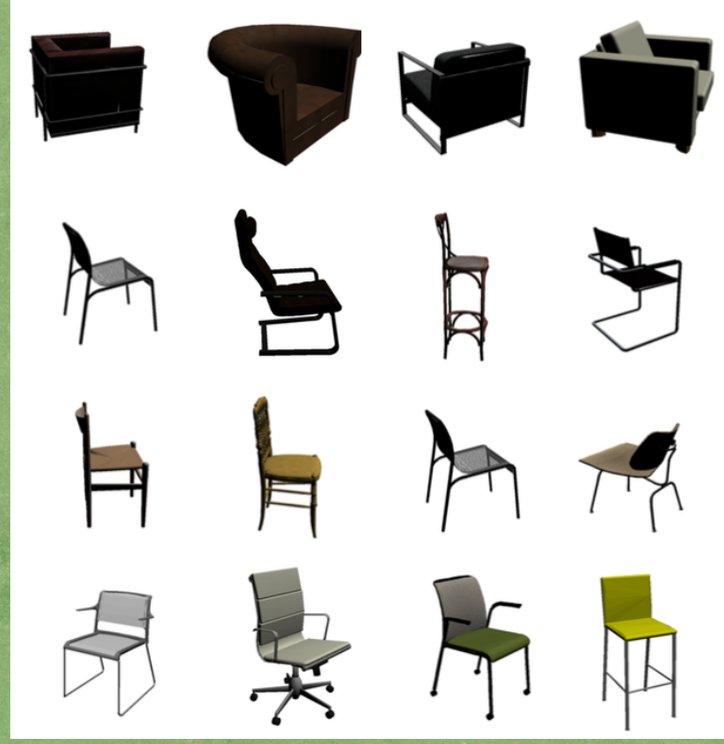
Qualitative Results (Synthetic images)

Questions

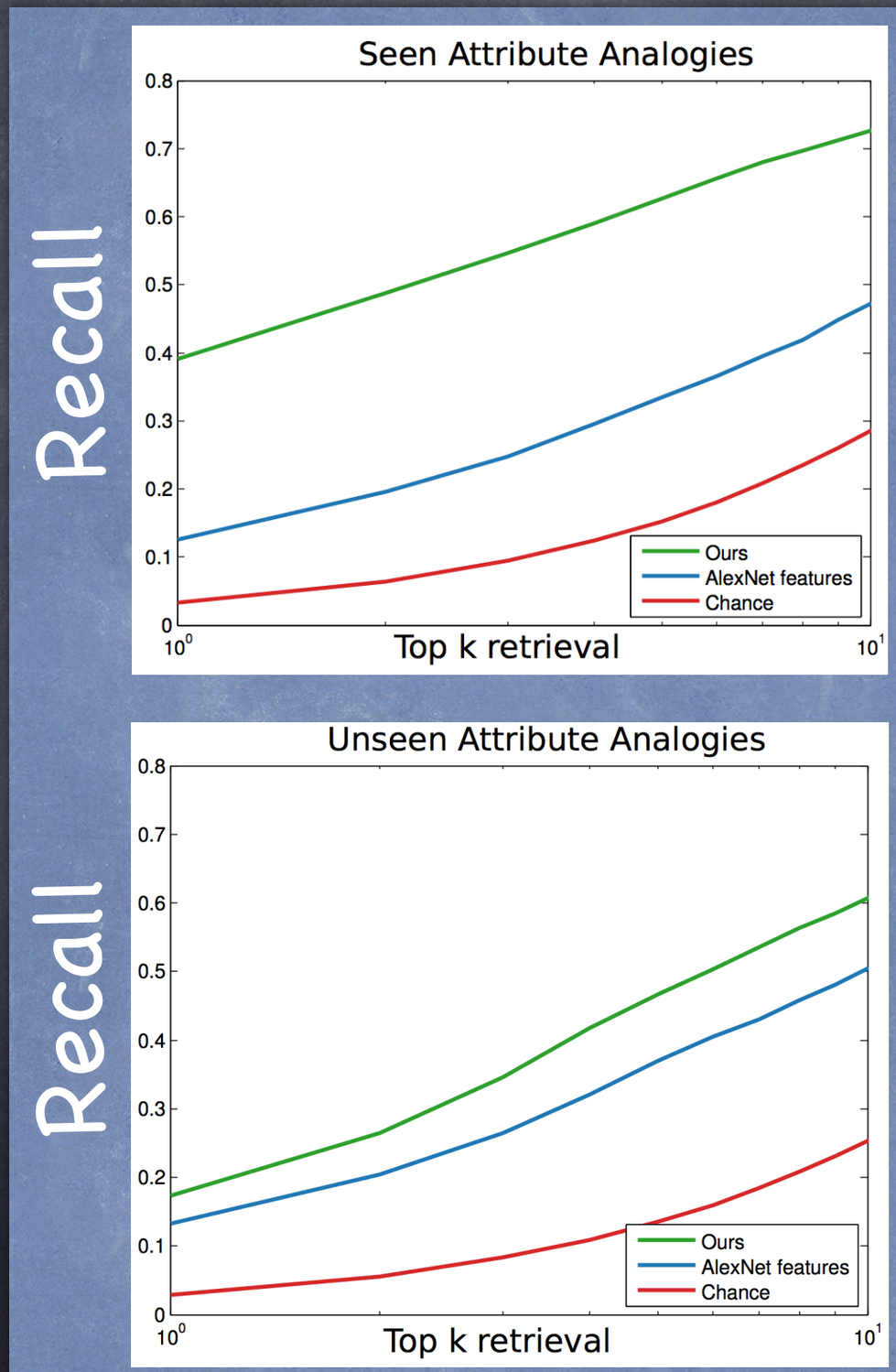


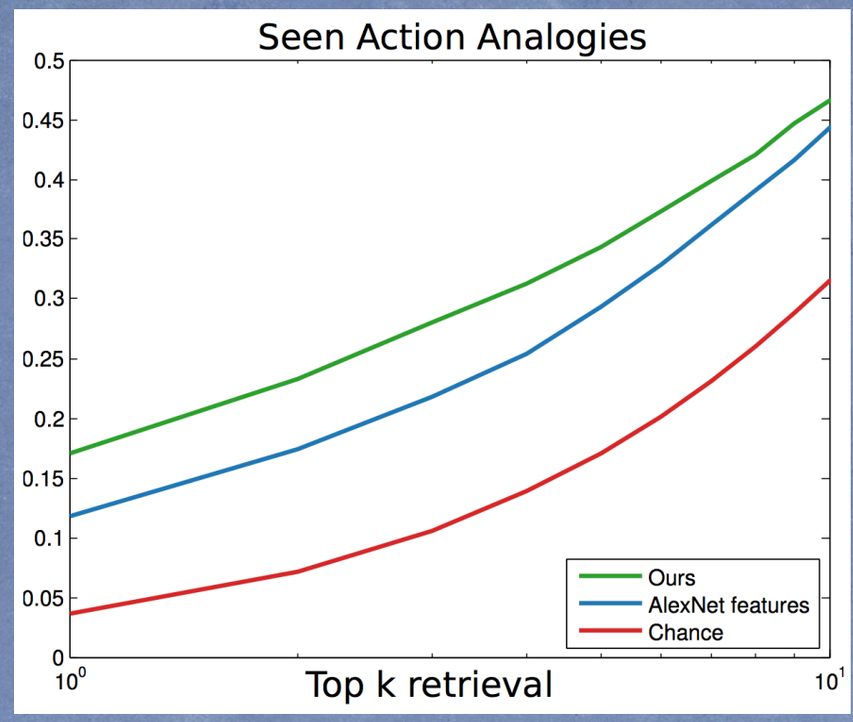
Ours

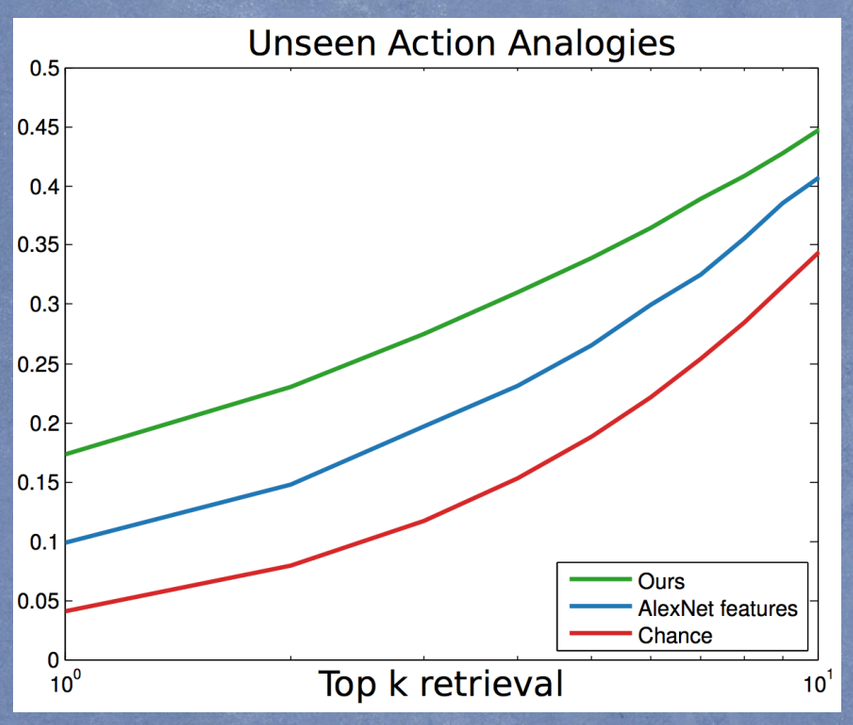
Baseline



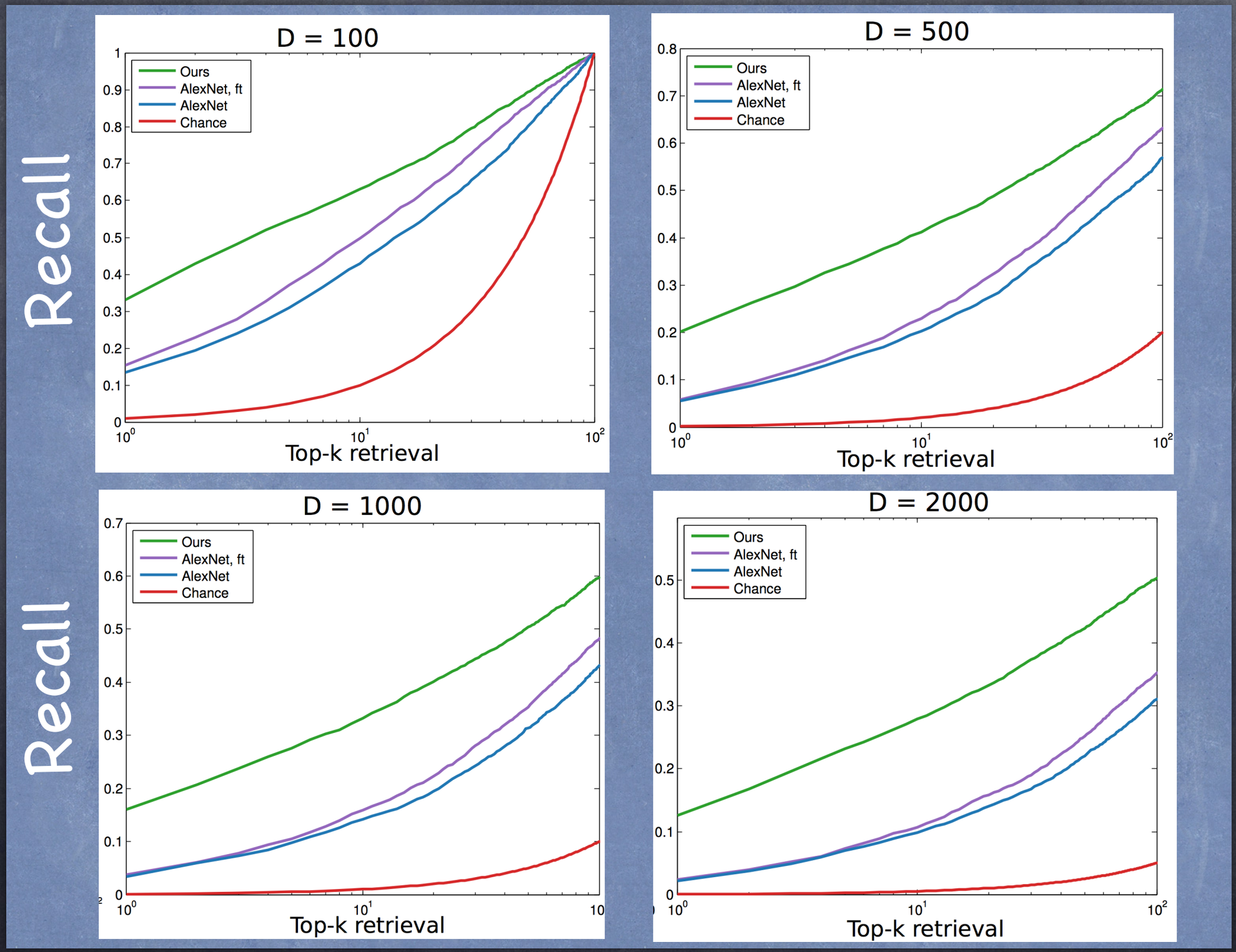
Quantitative Results (Natural images)







Quantitative Results (Synthetic images)



```
*:**::?:??
```