Summarizing Visual Data Using Bidirectional Similarity

Team Name - LTMG
Project ID - 12
Repository :
https://github.com/Computer-Vision-IIITH-2021/project-lmtg

Lakshmi Madhuri - 2018101116
Mohit pavan kumar - 2018102016
Tejaswini Anuhya Suma - 2018102018
Gowthami Gongati - 2018102048

Outline of The Paper-(1/3)

- We propose a principled approach to summarization of visual data (images or video) i.e storing the “important” visual content in a smaller image generated exploiting the data redundancy.
- This can be done based on optimization of a well-defined similarity measure. A good “visual summary” should contain as much as possible visual information from the input data and also shouldn't not introduce new visual artifacts that were not in the input data
- We have an objective function which has constraints as follows:
  - Two signals S and T are considered visually similar if all patches of S (at multiple scales) are contained in T, and vice versa.

Outline of The Paper(2/3)

- The objective function is as follows:

$$d(S,T) = \frac{1}{N_S}\sum_{P\subset S}\min_{Q\subset T} D(P,Q) + \frac{1}{N_T}\sum_{Q\subset T}\min_{P\subset S} D(Q,P)$$

- Here S,T are two signals and P,Q are subsets of S and T respectively which are considered as patches.
- We are calculating Sum of Squared distances between every 2 patches P,Q

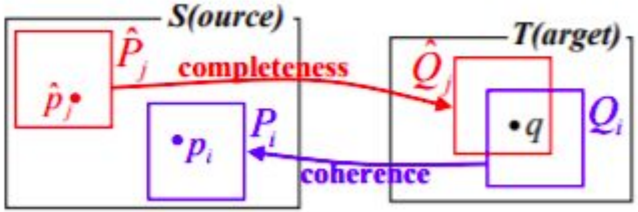
Outline of the Paper(3/3)

- The algorithm is as follows:-
  - There are 2 aspects to this algorithm-Completeness and coherence.
  - Assume S,T are 2 signals, if all patches of S(Source) are in T then T is complete even though T has new aspects introduced (if any)
  - Similarly if all patches in T are present in S without new patches being introduced then T is coherent.
- Our algorithm deals with image summarization such that the output image is both complete and coherent.

Method Overview(1/3):-

- Let $q \in T$ be a pixel in T, and let $T(q)$ denote its color.
- We calculate distance, d(S(source), T(target)) by adding both coherence term and complete term in it.
- We minimize this distance as follows:
  - Error every pixel $q \in T$ contributes to coherence term and complete term can be calculated as follows:-
    - Let $Q1, \ldots, Qm$ denote all the patches in T that contain pixel q and $P1, \ldots, Pm$ denote the corresponding (most similar) patches in S (i.e., $Pi = \arg\min_{P \subset S} D(P, Qi)$).
    - Let $p1, \ldots, pm$ be the pixels in $P1, \ldots, Pm$ corresponding to the position of pixel q within $Q1, \ldots, Qm$.

Method Overview-(2/3)



$$\frac{1}{N_S}\sum_{j=1}^{n}(S(\hat{p}_j) - T(q))^2$$ -the coherence term

$$\frac{1}{N_T}\sum_{i=1}^{m}(S(p_i) - T(q))^2.$$ -complete term

Error or dissimilarity measure would be sum of these two. We differentiate this error term to find T(q) which minimizes this expression to get

$$T(q) = \frac{\frac{1}{N_S}\sum_{j=1}^{n} S(\hat{p}_j) + \frac{1}{N_T}\sum_{i=1}^{m} S(p_i)}{\frac{n}{N_S} + \frac{m}{N_T}}$$

Method Overview-(3/3)

The above can be achieved as follows:-

Given the target image T (l) obtained in the l-th iteration, we compute the colors of the target image T (l+1) as follows:

1. For each target patch Q ⊂ T (l) find the most similar source patch P ⊂ S (minimize D(P, Q)). Colors of pixels in P are votes for pixels in Q with weight 1/NT .

2. In the opposite direction: for each Pˆ ⊂ S find the most similar Qˆ ⊂ T (l) . Pixels in Pˆ vote for pixels in Qˆ with weight 1/NS.

3. For each target pixel q take weighted average of all its votes as its new color T (l+1)(q). (Color votes S(pi) are found in step 1, S(ˆpi) in step 2.)

We iterate till we get a convergence point which would be output of image

summary

Goals-To perform Image Summarization

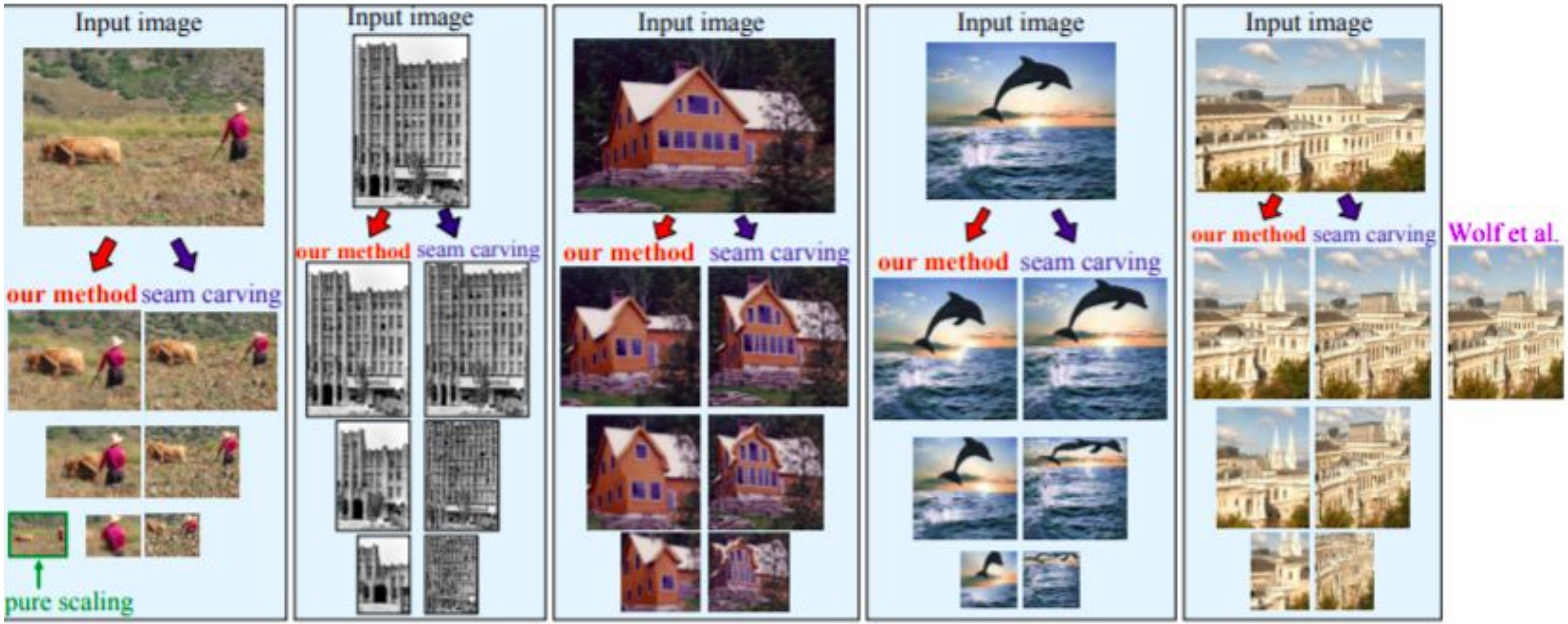


Figure 6. **Image summarization results.** *Our method exploits redundancies in the images (bushes, waves, windows of the buildings, etc.), often creating coherently-looking images even for extremely small target sizes. "Seam Carving" prefers to remove low-gradient pixels, thus distorts the image considerably at small sizes, when there are no more low-gradient pixels left. Please view video on demo website. The Dolphin image is from Avidan and Shamir [1], the right-most building image from Wolf* et al. *[16].*

Applications of this approach

- Image/video summarization
- Data completion and removal
- Image synthesis
- Image collage
- Photo reshuffling
- Auto-cropping

Image montage result

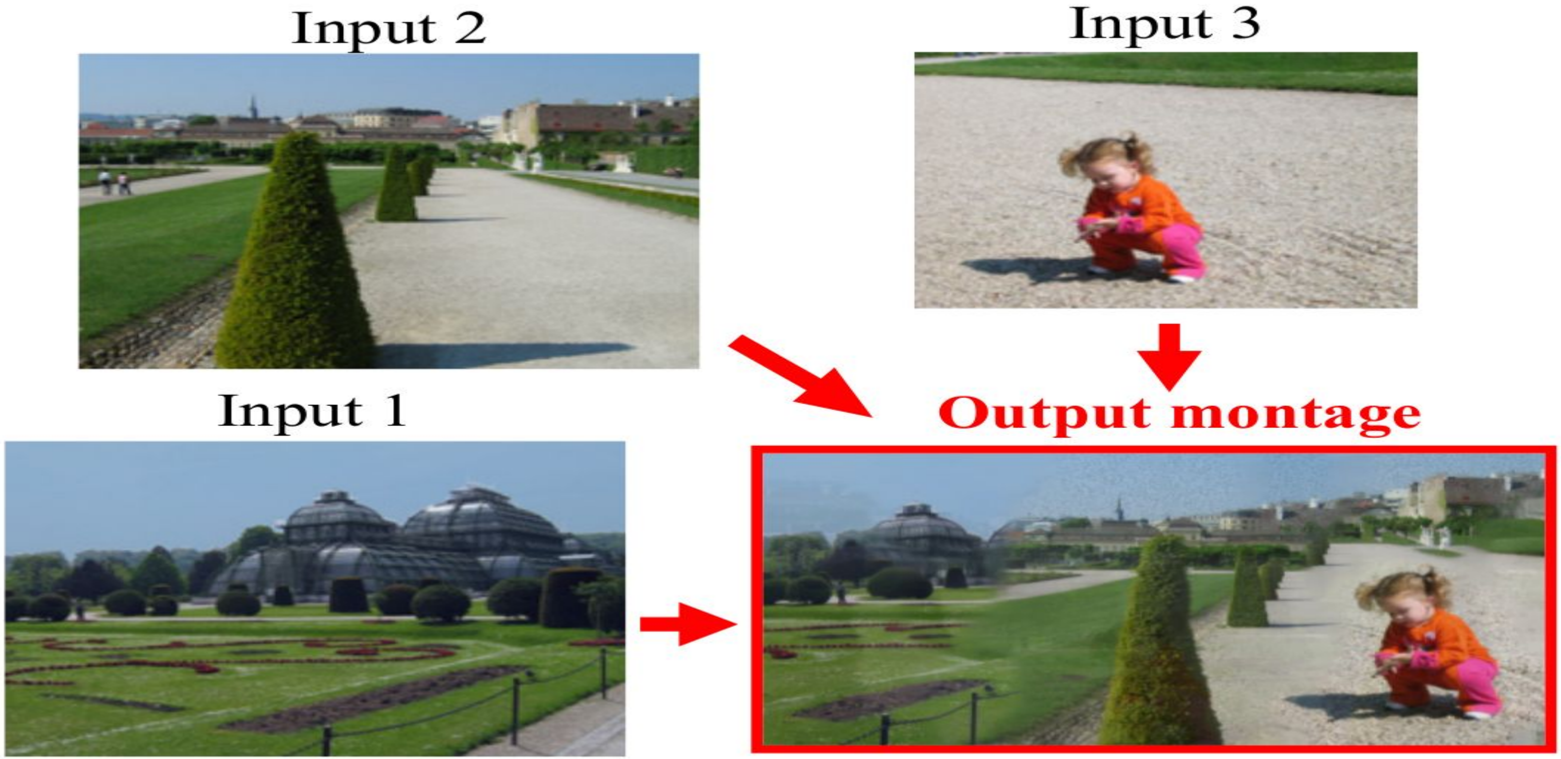
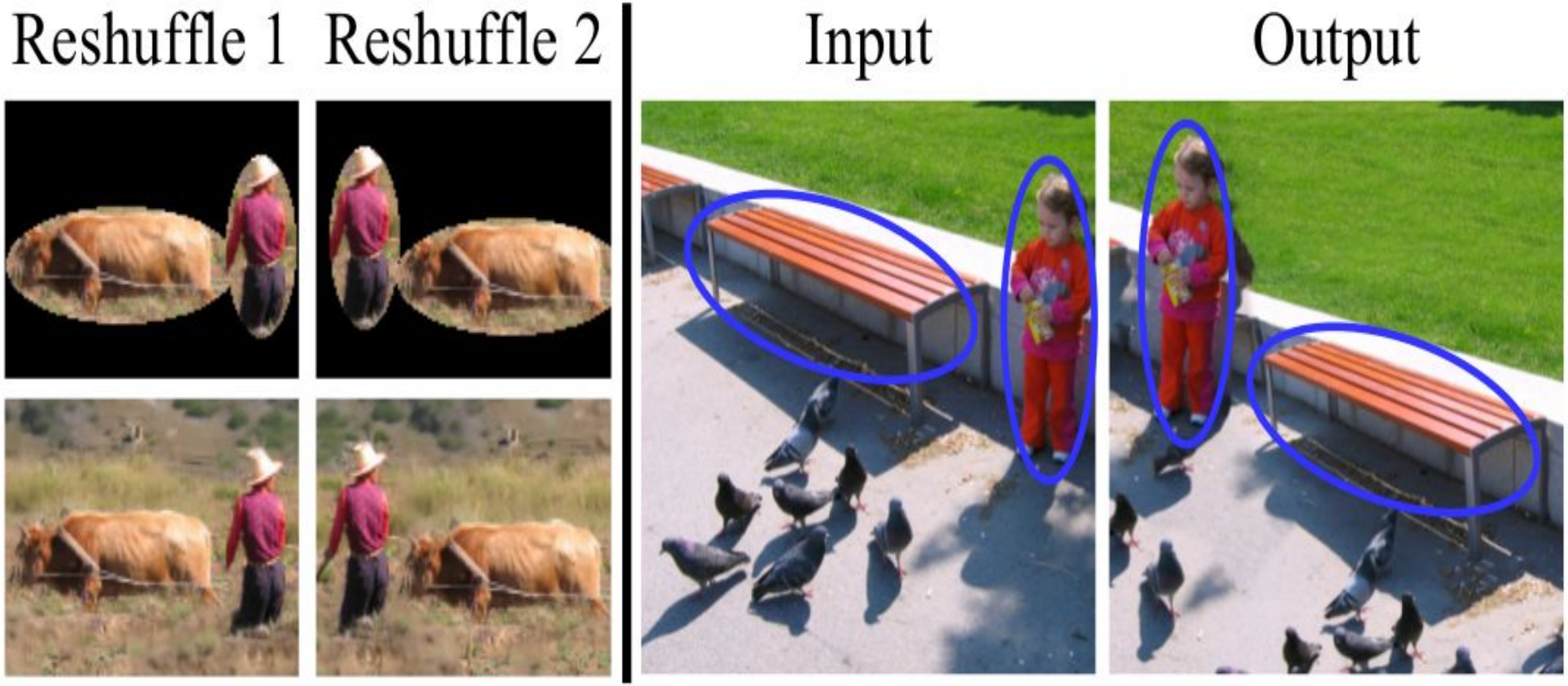
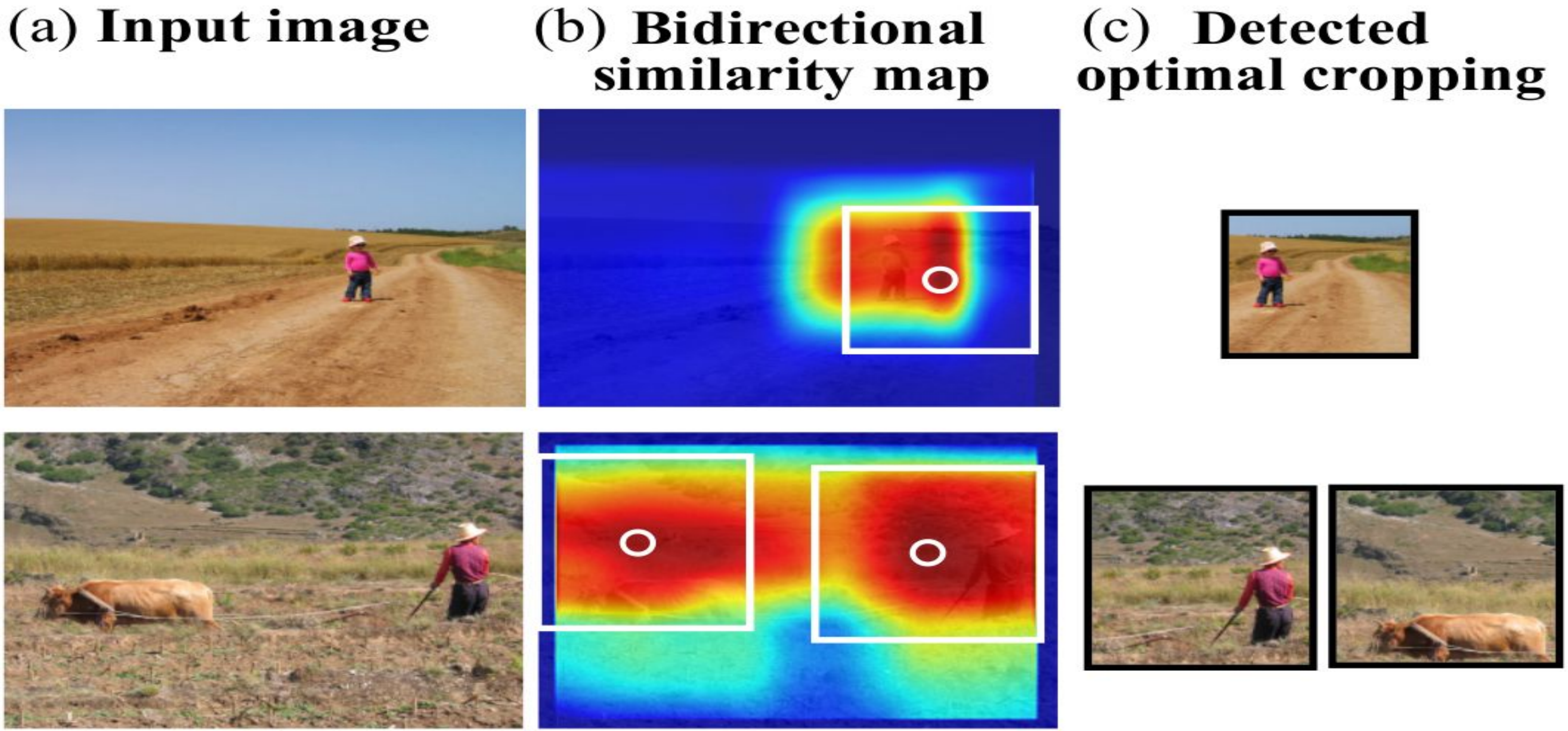


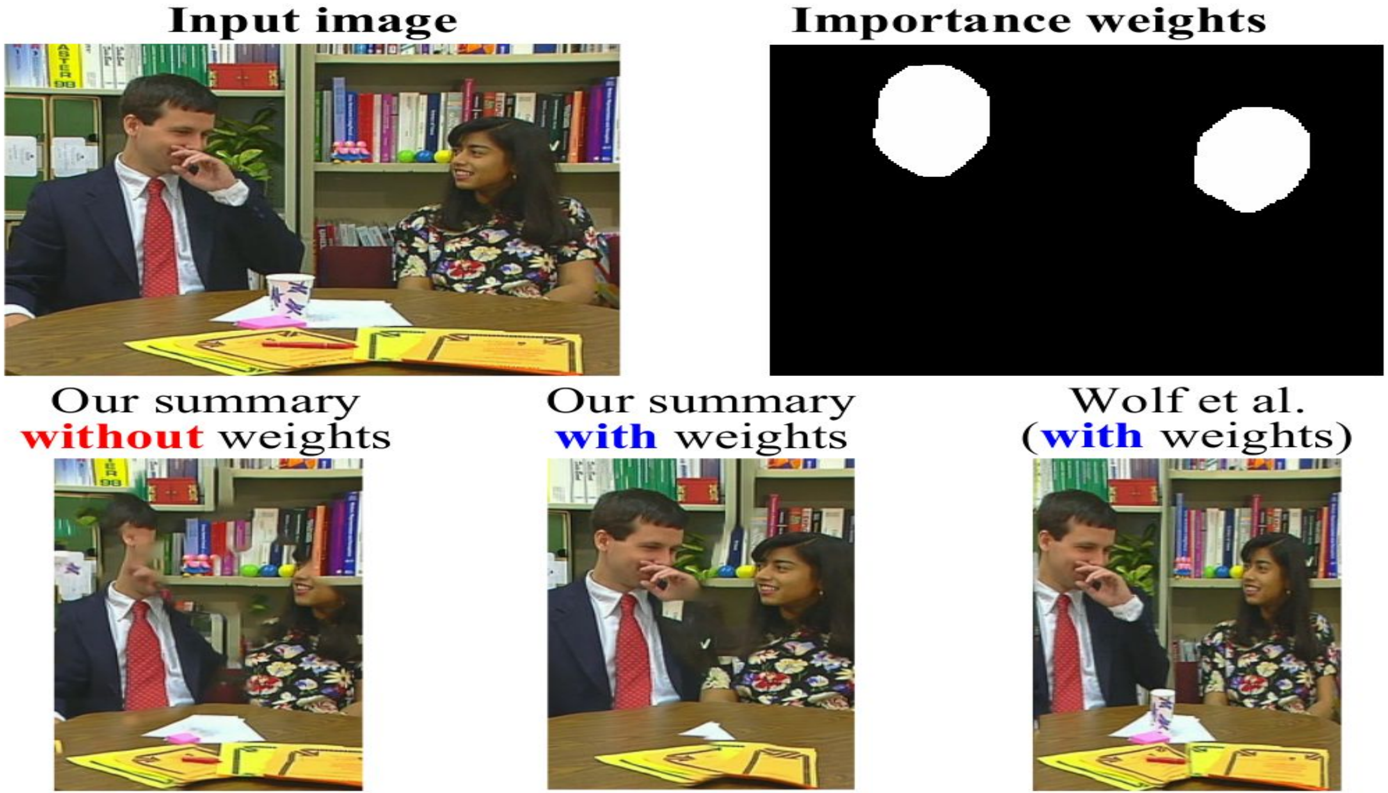
Photo reshuffling



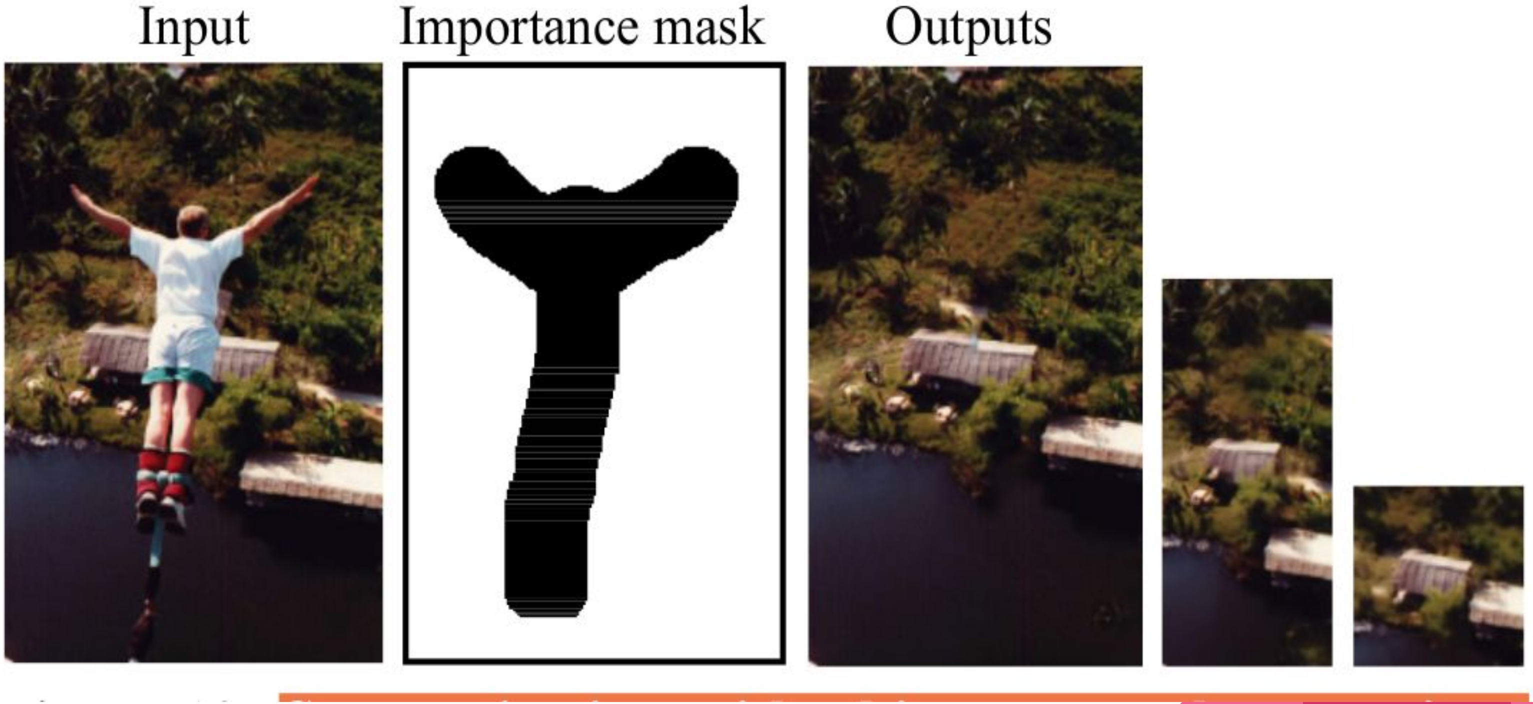
Automatic optimal cropping



Incorporating non-uniform importance



Summarization with object removal constraints



Conclusion

- ➢ Includes a bidirectional similarity measure between two images/videos of different sizes.
- ➢ Described a principled approach to retargeting and summarization of visual data (images and videos) by optimizing this bidirectional similarity measure.
- ➢ And the applications of this approach to image/video summarization, data completion and removal, image synthesis, image collage, photo reshuffling and auto-cropping.

Timeline

Understanding paper in depth and begin to implement paper	25 February
Basic implementation of distance function and debugging	10 March
Rough implementation of the whole algorithm	10 March - 25 March
Final implementation	13 April
Final testing	14 April
Documentation	15 April
Final commit	17 April (approximately)

Deliverables for the mid evaluation, We will going to complete the work up-to the implementation of bidirectional similarity algorithm and present some outputs of how our approach will work on the different applications listed above.