November 2019

Automated image retargeting at scale using a generative adversarial network

Jingbin Wang

Ramin Zabih

Chen Wang

Richard Strong Bowen

Michael Wilber

Automated image retargeting at scale using a generative adversarial network

ABSTRACT

Given the diverse variety of device screens, approaches that rely on manual creation of various forms of an image for each display situation do not scale. Therefore, automated techniques are required to process and resize images to adjust for different devices on which the image is to be displayed. The goal of automated retargeting is to change the aspect ratio of the original image while preserving the semantic and visual meaning of the content within the image. Currently, typical retargeting processes involve a multitude of operations, use a variety of heuristics that are tuned manually, and do not work effectively for all images. This disclosure describes a generative adversarial network (GAN) used to retarget an input image to a different aspect ratio. The described approach involves scaling each pixel row within an image by a factor between 0 and 1.

KEYWORDS

- Image content
- Image retargeting
- Generative Adversarial Network (GAN)
- Differentiable retargeting layer
- Image scaling
- Advertising image
- Image elements
- Image semantics
- Content preserving deformation

BACKGROUND

In many situations, an image within given web page or app needs to be displayed on device screens that differ in terms of their properties, such as size, resolution, aspect ratio, etc. For instance, mobile devices usually have smaller screens that traditional desktop or laptop computers. Owing to the differences in the properties of the screen on which content is being displayed, it may not be feasible or optimal to show the original image unchanged. Given the diverse variety of device screens, approaches that rely on manual creation of various forms of the image for each situation do not scale.

Retargeting refers to automated techniques to modify an image to make it fit within the constraints imposed by display parameters. The goal of retargeting is to change the aspect ratio of the original image to a ratio that is optimal to display the image within the overall web page or app based on the properties of the screen on which the image is to be displayed, while preserving the semantic and visual meaning of the content within the image. In essence, the retargeting process transforms the input image to create new versions at different sizes while attempting to maintain the meaningfulness of the content within the image. For instance, retargeting can be used to display images of advertisements that are placed at various locations within web pages or apps and need to be rendered on a variety of device screens. The ability to automatically generate optimized images via retargeting allows ad content to scale, thus increasing inventory and boosting the ability to serve a larger audience. This can increase the efficiency and revenue for advertising networks.

Current retargeting techniques involve a multitude of operations, such as seam carving, isotropic scaling, anisotropic scaling, etc. Further, for content such as faces, text, or logos, a variety of heuristics are used; such heuristics need to be tuned manually. Further, the retargeting process often does not yield a retargeted image at all or outputs an image at a size different than the one required for the context. Further, the processes can produce retargeted images that introduce content distortions that alter or degrade the semantic meaning of the content within the image. For example, retargeting an image of a plane can result in an input image that shows the plane with a disproportionately tall tail wing, a retargeted image of a logo could distort or leave out critical elements within the logo, etc.

DESCRIPTION

This disclosure describes a generative adversarial network (GAN) that can be used to retarget an input image. The GAN can be used to shrink the height of an image while keeping its width the same as the original, thus providing the ability to change the aspect ratio.

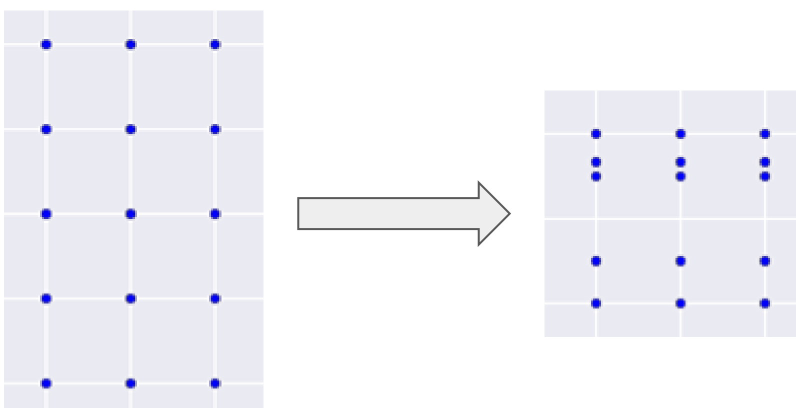

Fig. 1: Shrinking from height 5 to height 3

Fig. 1 illustrates resizing an image from height 5 to height 3. The illustrated image has 3 columns and 5 rows. In retargeting an image using a GAN as described herein, each row of pixels in the image is assigned a real number between 0 and 1, where the number represents the amount of scaling to be applied to the row when generating the retargeted image. A row of pixels is omitted from the retargeted image when the number is assigned the number value of 0 and is copied unchanged when it is assigned a number value of 1.

In contrast to approaches that make a binary decision on whether each row as included or excluded, the described techniques perform scaling of each row that is assigned a number other than 0 or 1 in proportion to the assigned number and including the scaled output in the retargeted image. Thus, each seam within the image has a continuous height and is a straight horizontal line. Use of constant height rows ensures that there is no waviness in the output image. Further, X coordinates are not modified.

The deep retargeting GAN as described above is obtained using end-to-end training without involving any pre-trained networks. Training data for the network can include tall images, e.g., 300x600 pixels, and short images, e.g., 300x250 pixels. The network is trained such that scaling factors assigned to each pixel row within the input image take into account aspects such as white space, likely focus(es) of user attention on content within the image, etc.

The generator can be implemented as a residual convolutional U-network with a differentiable resize layer. The generator network determines the scaling factors that are used to resize the image to a target height while keeping the width constant. In the resizing step, the generator can obtain a row height per pixel of each column by solving an ordinary differential equation (ODE), with a boundary condition that the sum of row heights be equal to the target height. To enforce consistency across columns, the entire row is replaced with a weighted average of that row. Alternatively, a minimum value for each row (or other suitable values) can be used.

The discriminator can be implemented as a convolutional, residual downsampling network with batch normalization. The last layer of the discriminator can be a sigmoid or average pooling layer. The exact number of layers used is a hyperparameter.

The quality of images produced by the generator network is measured by analyzing the image via the discriminator to determine whether image produced by the generator can be distinguished from original images, e.g., the probability that a given image is an original image or a retargeted image.

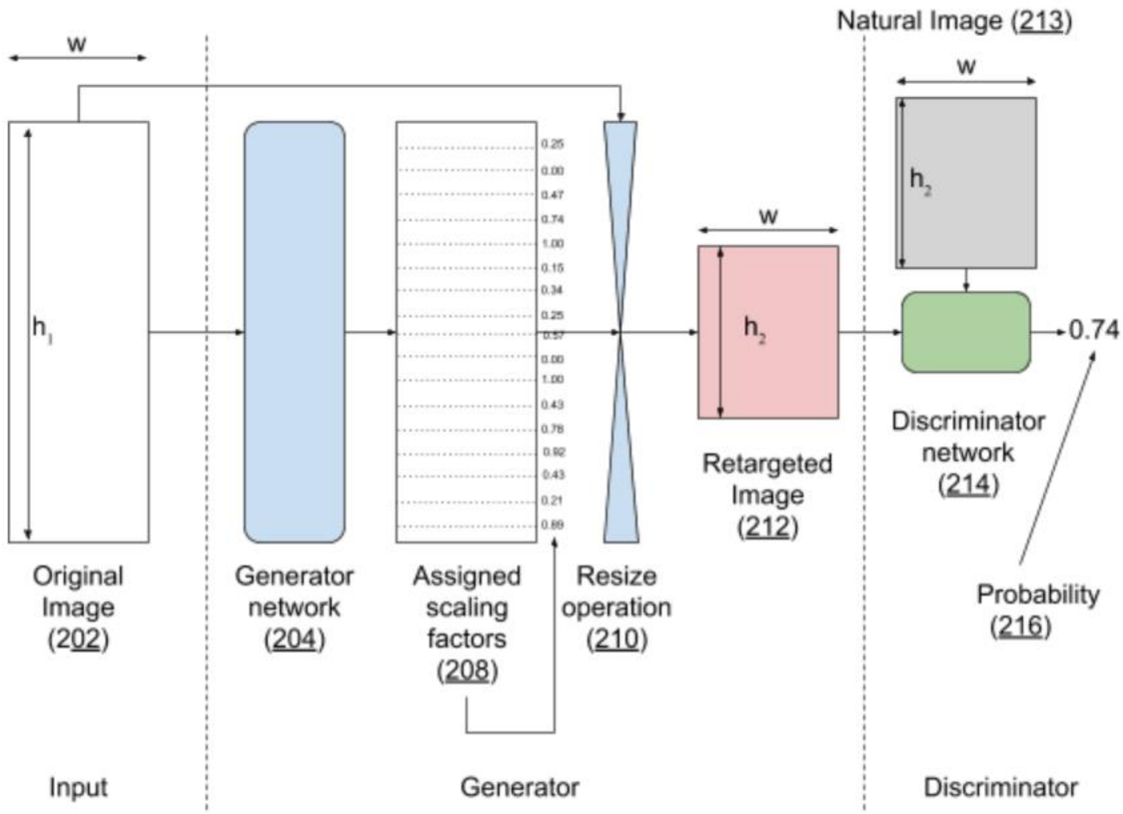

Fig. 2: Generative Adversarial Network retargeting an image to shorter height

Fig. 2 shows a configuration of a generative adversarial network to retarget images. Original images (202) of height h_1 and width w are retargeted to a height h_2 while keeping the width constant. A generator network (204) is used to process each original image to determine scaling factors (208) for each row of pixels within the image. Each scaling factor is a real number between 0 and 1, both inclusive. A resize operation (210) is performed using the scaling factors to generate a retargeted image (212) by scaling each row within the original image by the corresponding scaling factor. Rows with a scaling factor of 0 are omitted completely from the retargeted image while those with a scaling factor of 1 are retained in their original form. The retargeted image resulting from the above process is evaluated via a discriminator network (214) to output a probability (216) that the retargeted image is indistinguishable from original images provided to the discriminator.

During a training stage, the discriminator network is provided images resized by the generator network (e.g., of size h_2*w) as well as natural images (213) that have not been resized (e.g., also of size h_2*w). The goal of the training is to obtain images from the generator network that a trained discriminator network cannot differentiate from natural images. Once trained in this manner, the generator network can be used to retarget input images. Training can be carried out using a large set of images that are a representative sample of the kind of images that the generator network is utilized to retarget, e.g., advertising images.

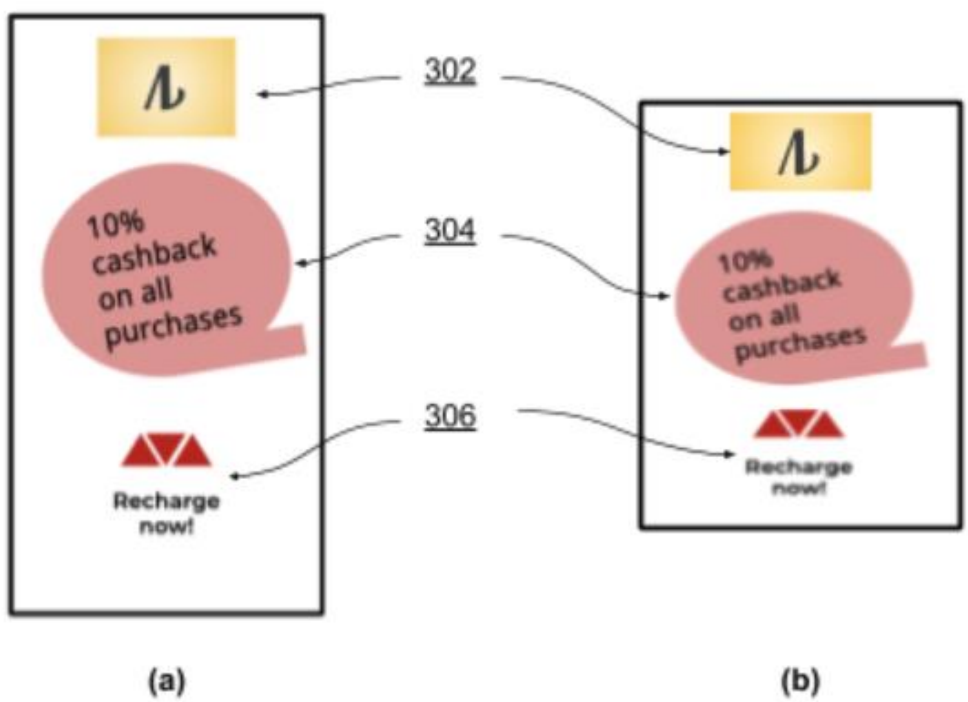

Fig. 3: (a) Original image; (b) Retargeted image

Fig. 3(a) illustrates an example original image and Fig. 3(b) illustrates a corresponding retargeted image. The images include three portions - a logo (302), text on top of a background image (304), and a second logo that includes text (306). As can be seen, the width of both images is identical, while the retargeted image has less height than the original image. In the retargeting, the height of different rows of pixels of the original image is adjusted differently. For example, whitespace near the top and bottom of the image - above 302 and below 304 as well as between the elements 302, 304, and 306 is reduced, e.g., these rows may be assigned a numeric value of 0. Further, the height of various elements is adjusted differently, e.g., logo 306 has a more pronounced reduction in height compared to other portions of the image.

The techniques described above involve end-to-end learning of attention metrics with fairly simple approaches, such as texture, entropy, gradient magnitude, etc. The operation can be enhanced by the addition of other approaches for attention. More flexibility can be achieved by varying x as well based on saliency of different parts of the image , e.g., to cause certain parts of the image to grow or shrink, e.g., content-portions such as logos or text can grow, while blank/background space can shrink during resizing. Further, data from rigorous manual comparisons, large scale A/B testing, and automated quality metrics such as text size measurements can be incorporated to improve the generator network.

The described techniques can also be improved by partitioning the original image into various semantic elements, such as logos, photos, text, etc. and inferring relationships between these elements. For example, an image advertising a washing machine may contain the photo a washing machine, a logo of the manufacturer, text describing the make and model of the machine, etc. Based on the target output size and placement for the retargeted image, each of the elements within the image can be processed separately to yield a corresponding retargeted image.

Such individually retargeted images of the elements can then be arranged in a variety of ways to generate a suitable combined retargeted image that meets the given display constraints. For instance, such an operation can be effective for converting a large width image of a shorter height within a tall and narrow content space by rearranging the elements within the image such that all of the content of the original is maintained with reduced distortion.

Application of the techniques described in this disclosure can enable image resizing at scale with minimal to no distortion of the semantic content of the image, thus enhancing the user experience of viewing image content on devices with varying screen capabilities. In case of advertising images, retargeting automatically using a generator network as described herein can boost the effectiveness and comprehension of the advertisement independent of the size at which the advertisement is displayed. As such, retargeting can translate into higher user engagement with the advertisement.

CONCLUSION

This disclosure describes a generative adversarial network (GAN) used to retarget an input image to a different aspect ratio. The described approach involves scaling each pixel row within an image by a factor between 0 and 1. The scaling factors take into account aspects such as white space, likely focus of user attention on content within the image, etc. The generative network can also be implemented to support more sophisticated retargeting modes such as scaling in both dimensions, prioritizing the preservation of specific content, etc. Further, the input image can be separated into various semantic elements and relationships between the elements can be inferred. Each element can then be retargeted separately and individually retargeted images can be arranged to generate a suitable combined retargeted image. The techniques described in this disclosure can enable image resizing at scale with minimal distortion the semantic content of the image and can be used to improve delivery of advertising images.

REFERENCES

1. Cho, Donghyeon, Jinsun Park, Tae-Hyun Oh, Yu-Wing Tai, and In So Kweon. "Weakly-and self-supervised learning for content-aware deep image retargeting." In Proceedings of the IEEE International Conference on Computer Vision, pp. 4558-4567. 2017.