



HAL
open science

Creative Intelligence – Automating Car Design Studio with Generative Adversarial Networks (GAN)

Sreedhar Radhakrishnan, Varun Bharadwaj, Varun Manjunath,
Ramamoorthy Srinath

► **To cite this version:**

Sreedhar Radhakrishnan, Varun Bharadwaj, Varun Manjunath, Ramamoorthy Srinath. Creative Intelligence – Automating Car Design Studio with Generative Adversarial Networks (GAN). 2nd International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE), Aug 2018, Hamburg, Germany. pp.160-175, 10.1007/978-3-319-99740-7_11 . hal-02060050

HAL Id: hal-02060050

<https://inria.hal.science/hal-02060050>

Submitted on 7 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Creative Intelligence – Automating Car Design Studio with Generative Adversarial Networks (GAN)

Sreedhar Radhakrishnan*^[0000-0002-4098-103X], Varun Bharadwaj*^[0000-0001-8986-9682],
Varun Manjunath*^[0000-0002-7769-160X] and Ramamoorthy Srinath^[0000-0003-1818-8577]

PES University, Bengaluru 560085, India

sreedhar1895@gmail.com, varunbharadwaj1995@gmail.com,
varunmanjunath2012@gmail.com, ramamoorthysrinath@gmail.com

Abstract. In this paper, we propose and implement a system based on Generative Adversarial Networks (GANs), to create novel car designs from a minimal design studio sketch. A key component of our architecture is a novel convolutional filter layer, that produces sketches similar to those drawn by designers during rapid prototyping. The sketches produced are more aesthetic than the ones from standard edge detection filters or gradient operations. In addition, we show that our system is able to generate hitherto unseen perspectives of a car, given a sketch of the car at just a single viewing angle. For extensive training, testing and validation of our system, we have developed a comprehensive, paired dataset of around 100,000 car images (with transparent backgrounds) and their respective sketches. Our work augments human intelligence and creativity using machine learning and deep neural networks. Our system has the significant benefit of reducing the cycle time in the sketch-to-image process which has largely been considered a creative domain. This is achieved by learning to interpret a preliminary sketch drawn by a designer, to generate novel visual designs in a matter of seconds, which may otherwise require considerable time and effort. While the system enhances the productivity of the designer, the machine learning enhanced design visualizations can cut costs during the product prototyping stage. Our system exhibits good impactful potential for the automobile industry and can be easily adapted to industries which require creative intelligence.

Keywords: Computational Creativity, Generative Adversarial Networks, Automobile Design, Deep Learning, Computer Vision, Sketching Filter.

1 Introduction

‘A picture is worth a thousand words’, is an idiom that resonates with visual designers across industries. It is extremely relevant today, considering the scale at which product design has engendered innovation in the last decade. One of the most important phases in product design and manufacturing is prototype development. These visualizations are the best way to communicate ideas beyond language barriers. Developing multiple novel prototypes for stakeholders is an arduous task and one that

requires an amalgamation of creativity and tenacity. A designer first channels his creativity in the form of a ‘napkin sketch’, which is a rudimentary sketch that captures the thought process prior to developing prototypes. Although developing a napkin sketch may not take very long, developing multiple novel prototypes is a very time-consuming task. In today’s fast-paced environment, every minute saved would result in reduced cost for an organization. One such industry where innovation in product design results in high quality products and revenue is the automobile industry. Creating a palette of designs before zeroing in on the right design prior to manufacturing is a creative and expensive bottleneck in the production process.

In this paper, we seek to aid artists and car designers by breaking the creative bottleneck during car sketch-to-design translation. Empirical studies and extensive research such as ‘Role of sketching in conceptual design of car styling’ [1] have shown the primary importance of car sketches in both cost reduction and expediting the idea-to-product process.

We propose a system that uses GANs to augment the creativity of designers by generating multiple novel car designs in a matter of seconds. By leveraging semantic information of a given studio sketch, our system also generates car designs in multiple hitherto unseen perspectives. In addition, we show interesting results where a model trained only on car images shows serendipitous and visually appealing results for sketches of bikes and even non-automobile entities such as trees.

1.1 Generative Adversarial Networks

The domain of deep learning for computer vision has witnessed rapid progress since 1998, where the work on ‘Gradient-Based Learning Applied to Document Recognition’ [2] discussed Convolutional Neural Networks (CNN). AlexNet [3], that won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) further led to prodigious amount of research in CNN. AlexNet architecture, marked the first time when a CNN achieved a top five test error rate (rate at which, given an image, the model does not output the correct label with its top five predictions) of 15.4 percent.

Generative Adversarial Networks (GANs) [4], a class of machine learning algorithms used in unsupervised learning, introduced in 2014 have shown immense potential in areas of domain adaptation and image translation. GANs introduced a powerful concept where there are two adversarial models, a generator and a discriminator, with the models being trained using a min-max method. In other words, the generator produces images by capturing a data distribution while trying to minimize the classification accuracy of the discriminator. Essentially, it is the generator’s motive to deceive the discriminator while the discriminator tries to maximize its classification accuracy. The discriminator’s role is to classify an input image as an original image or one that is not (generated from the generator). Mathematically, the value function for a discriminator and a generator is given by a simple formulation of cross-entropy and expected values.

The adversarial networks are trained until the generator can satisfactorily produce outputs that the discriminator classifies as real. Although stability and stopping

criterion were major concerns, GANs fostered research in the areas of image translation, domain adaptation and various forms of generative tasks.

There has been a lot of work on improving the stability of GANs such as Deep Convolutional Generative Adversarial Networks (DCGANs) [5] and Wasserstein GAN (WGAN) [6]. Coupled Generative Adversarial Networks (CoGAN) [7] introduced an architecture where two GANs are used to learn a joint distribution of multi-domain images. This is achieved through weight sharing in both the generator and the discriminator in the layers where high level features are extracted. Conditional Adversarial Networks [8] not only learn a mapping from input to output domain but also learn a loss function to train this mapping. Conditional Adversarial Networks, thus expect a paired dataset for training. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks (Cycle GAN) [9] employs an inverse mapping such that the pairing is obtained only if an image X_i in domain X produces an image Y_i in domain Y such that translation of Y_i from domain Y back to domain X produces X_i . This mapping is useful in cases when a paired dataset is not available.

While Conditional Adversarial Networks perform well in image translation tasks, the architecture expects a paired dataset, which in most cases either does not exist or is difficult to acquire. Hence, we developed a paired dataset of around 100,000 car images (with a transparent background) along with their respective sketches. The dataset will not only help us leverage Conditional GANs in our system but will also be an aid for researchers worldwide supporting further advancement and innovation in related fields. Researchers can contact the authors of this paper to obtain the dataset. The dataset will also be open-sourced for public domain access.

1.2 Related work

Computational creativity and its related fields have attracted a lot of attention in the research community. There has been work in developing a standard pipeline for sketch based modelling [10] which encompasses methods and techniques to enable users to interact with a computer program through sketching. The three main stages are input, filter and interpret. Later, image acquisition and visual rules are used for preventing over-sketching and similar defects. Although the above paper does not use machine learning for modelling input sketches, the goal of improving human productivity remains the same.

Work on design generation in fashion and clothing industry, such as Visually-Aware Fashion Recommendation and Design with Generative Image Models [11] use feature representations from CNNs and Siamese neural networks to make a fashion recommendation system. Further, in unison with GANs, the same system is used in a generative manner, where new designs are fabricated conditioned on the given user profile and product category. However, it is not a sketch-based modelling system.

There has also been prior work in interpreting human sketches for image retrieval in large databases [12]. Humans can visualize and draw objects with a certain degree of accuracy. Thus, sketch modelling based image retrieval systems have the potential to perform better than traditional retrieval systems. While our work focuses on deep learning based rapid prototype generation from a sketch, human sketch based

information retrieval provides an intuitive interface for fast image retrieval reflecting human imagination.

A recent work presents an implementation where GANs are used in a user interactive program to generate realistic images constrained by a natural image manifold [13]. The object can be distorted and reshaped and the GANs automatically adjust the output keeping all edits as realistic as possible. The manifold of natural images is used as a constraint on the output of various image manipulation operations, this makes sure the result always lies in the learned manifold.

In our paper, we utilize GANs for learning the distribution across the training samples and creating new samples in the specific domain of automobile design, especially cars. For the training of conditional GANs, we fabricate artificial sketches using a novel convolutional-morphological filter.

2 Proposed System

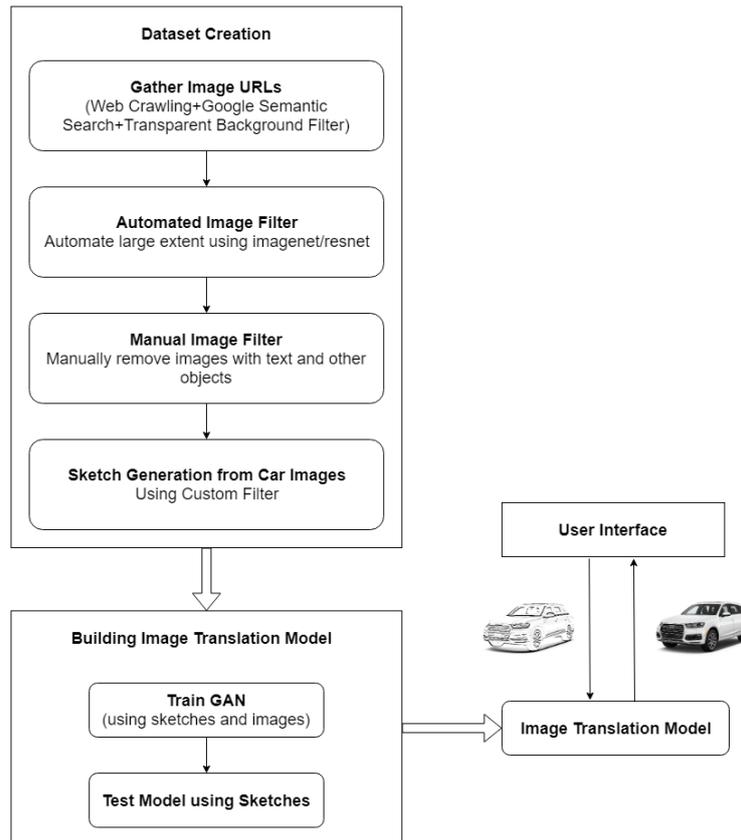


Fig. 1. Proposed GAN Based Car Sketch Image Translation System.

Our system, as shown in Fig. 1, essentially has three phases: dataset creation, machine learning model training and developing a user interface for the designer. Within the dataset creation phase, there are two core components, first being, the development of a comprehensive dataset of around 100,000 car images and the second being the generation of sketches using a new filter that we have engineered for the task.

In the model training phase, we have trained 12 machine learning models through which our system generates novel car designs covering various colour combinations and unseen car angles. The system also has a web based interface where the designer can upload sketches, select colours and perspectives and view the design visualizations.

3 Dataset



Fig. 2. Snapshot of sample images in the dataset.

The construction of a comprehensive dataset involved three main steps:

1. Identifying and mining transparent car images.
2. Machine learning based removal of images with abnormal artifacts (undesirable objects or symbols). This was implemented using ImageNet [14] and ResNet50 [15] for image recognition and for classifying and filtering out images with entities that are not cars.
3. The remaining images with abnormal artifacts in the dataset were eliminated through manual inspection, which required extensive effort. About 60% of the time in the dataset collection phase was spent on manual inspection.

The multiple colours, angles and consistent transparent backgrounds across all car images in the dataset ensured that the dataset formed the base of the system. The system expects high quality, clean data with no background inconsistencies or noise for ensuring successful training of the machine learning models. The developed dataset satisfies all the above conditions and thus forms a substrate for subsequent steps in developing our system. The quality and variety of images satisfies our requirements to get paired datasets of same and different perspectives of cars for training in the next phase of the pipeline. A snapshot of the car image dataset is shown in Fig. 2.

4 The BiSECT Sketching Filter

Generally, if an outline is required, the standard approaches are gradient operators or edge filters. Many famous algorithms appear in standard image processing and computer vision literature. Notable ones which produce appreciable results are the Marr-Hildreth filter, Laplacian operator, Difference of Gradients (DoG), Sobel filter, Scharr filter, Canny edge operator [16] and the Extended Difference of Gaussians (XDoG) [17]. Each of these have their own locality, sensitivity and orientation based advantages and disadvantages. However, for our use case, none of these standard techniques produced outputs that capture the artistic nuances of a human sketch.

Hence, our goal was to create a filter that produces sketches similar to those drawn by humans for training the system with a paired dataset. The filters were applied on car images such as Fig. 3 (a) and the outputs were analyzed. As shown in Fig. 3 (b), while the Canny edge detector is able to identify edges, it lacks the stroke detail of sketches. The output from XDoG, on the other hand, as shown in Fig. 3 (c), captures too many details, beyond what we expect in a simple sketch. The output of XDoG looks synthetic and does not have the sketch texture one would expect from a pencil drawn sketch.

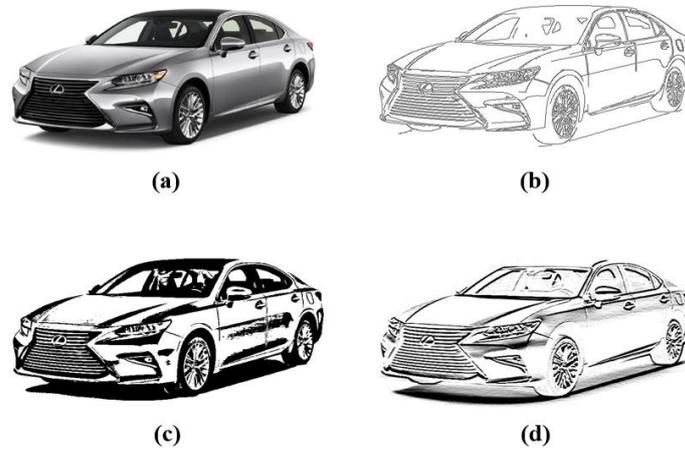


Fig. 3. Outputs of various filters. (a) shows the input image. (b), (c) and (d) show the outputs of the Canny, xDoG and BiSECT operators respectively.

There has been prior work on simulating sketches computationally such as ‘Automatic Generation of Pencil-sketch Like Drawings’ [18], using gradient estimation on smoothed images.

The sketching filter we propose is shown to perform well for car images, leading to a ‘napkin sketch’ kind of effect as shown in Fig. 3 (d). These sketches are similar to the preliminary drawings by designers or artists in the field of design and manufacturing.

We propose a new convolutional filter, the “BiSECT” sketching filter, which is an abbreviation for Bidirectional Sobel Enhanced with Closed Thresholds, which conveys the inner working of the filter concisely. The filter can be thought of as a “bisecting” operation that extracts the sketch artifacts from a given image.

4.1 Formulation of the BiSECT Sketching Filter

The BiSECT sketching filter was developed using a combination of existing standard convolutional kernels and morphological operations. A level of smoothing Gaussians and pixel-wise logical operations are used for combining and retaining the appropriate edges across different intermediate outputs. The filter was engineered as follows:

1. We use a combination of Sobel family operators because a single Sobel-Feldman convolutional filter restricts itself to finding edges in a particular orientation.
2. Combining the Sobel filter family outputs using a cascade of bitwise-OR and bitwise-AND operations allowed retaining important edges and having finer control over the final sketch.
3. After this series of convolution operations, we use a threshold layer to discard all the unnecessary levels of gray all over the intermediate outputs. We explored

Otsu's thresholding and adaptive thresholding techniques. After some experimentation, an appropriate global thresholding was satisfactory.

4. Finally, a 'morphological close' operation is performed on the thresholded output to coalesce all slightly disconnected true edges with structural elements of sizes $(8*1)$ and $(1*8)$ for both vertical and horizontal axes.

After aggregating and auditing our image dataset, each image from the 100,000 and odd car images was paired with its corresponding output from the above filter. This forms a standard paired dataset, which can be fed into conditional adversarial networks.

5 Experiments and Results

Experiments were run to train the system to generate novel designs by learning a pixel-level semantic mapping from sketches. The goal was to train the system to learn from minimal amount of data. This is because we wanted to adapt the system to the real world wherein the system will be continuously subjected to a plethora of sketches with variations.

The system learns the mapping from minimal data and generates high quality designs on significant amount of test data making the system robust and reliable. In the experiments below, we configured the model to train for 200 epochs on 200 paired images. The graphs in Fig. 4 and Fig. 5 show the generator and discriminator losses against the number of steps in training of the GANs.

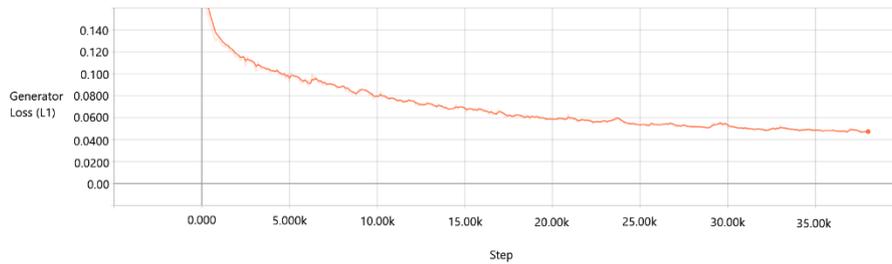


Fig. 4. Generator Training Loss Visualization



Fig. 5. Discriminator Training Loss Visualization

5.1 Single Colour Car Generation

We trained the system to generate designs in red, blue, white, and black as well as grayscale. We can observe that the system is able to generate state of the art car prototypes from preliminary sketches. The system correctly identifies regions in the sketch that need to be filled with a respective colour and texture. The system consistently is able to identify and augment even subtle details such as headlamp and tyre correctly.

Each of the input sketches visualize a diverse range of car models, shapes and sizes. The models correctly interpret the sketch, such as in Fig. 6 where the system has correctly distinguished between the skin and the seat of the convertible itself.



Input Sketch



Car Generated In Blue

Fig. 6. Translation of sketch to a blue prototype



Input Sketch



Car Generated In Black

Fig. 7. Translation of sketch to a black prototype

5.2 Mixed Colour Car Generation

In addition, we trained the system to generate prototypes consisting of an amalgamation of multiple colours. The combinations we trained the system on were

red and yellow, black and white, blue and gray as well as a four colour mix consisting of black, blue, reddish-brown and gray.



Input Sketch



Car Generated In Red and Yellow

Fig. 8. Dual colour output: Red and Yellow



Input Sketch



Car Generated In Blue and Gray

Fig. 9. Dual colour output: Blue and Gray

5.3 Multiple Design Visualizations

Our system enables designers to visualize multiple prototypes for a given sketch. Viewing multiple prototypes in juxtaposition allows the designer to make faster decisions regarding the final design. In addition, the system enables the designer to visualize or develop multiple prototypes quickly resulting in significant saving in cost, time and effort.

In the example shown in Fig. 10, we have generated three different car prototypes using grayscale, single color and multiple color combination models from a single sketch. While all the three models largely resemble each other structurally, each prototype has its own aesthetic appeal.

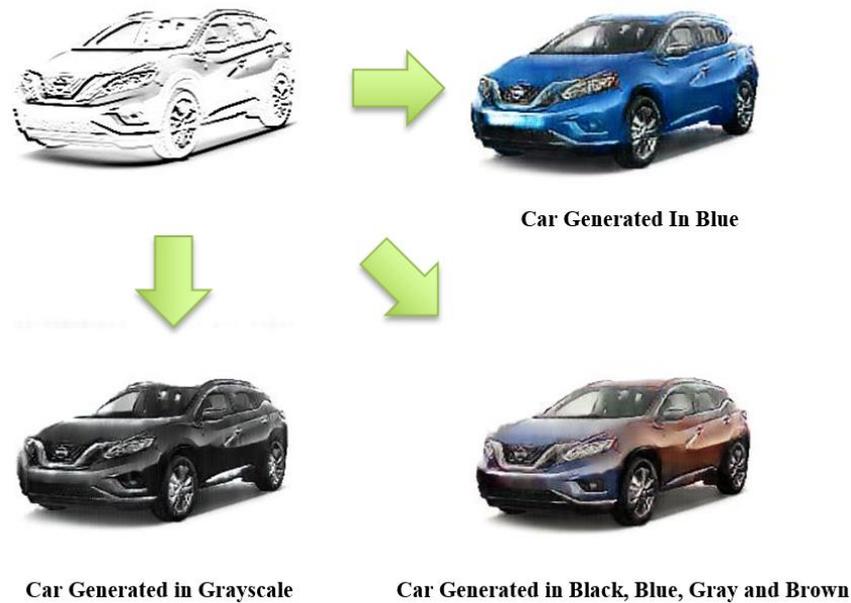


Fig. 10. Outputs from multiple models

With minimal training, the system is able to embellish the sketch to produce prototypes in multiple colours across various shapes and sizes, as shown in the figures in this section.

6 Artist Sketch Experiment

We obtained a shaded pencil sketch drawn by an artist. We used the GrabCut algorithm [19] to segment out the design. We then applied the “BiSECT” filter and tested the result on our machine learning models. The initial pencil shaded sketch image took considerable amount of time and our system interpreted the sketch and generated multiple prototypes in less than a minute.

In this particular case, the artist spent 8 hours spread over a period of 5 days in creating the initial sketch on paper. The sketch in Fig. 11 is evidently more detailed than a napkin sketch. Our system is capable of processing sketches that do not require extensive amount of time and effort to draw. Thus, with a preliminary sketch such as those shown in section 5, our system will generate multiple designs and the artist need not spend a long time to add extensive details.



Pencil Shaded Sketch



Generated Blue Prototype



Generated Black Prototype

Fig. 11. Testing on an artist's sketch

7 Generating Hitherto Unseen Perspectives of a Car

We have made some interesting headway in using the system to predict and generate how a car will look from various hitherto unseen angles and perspectives. This is done purely based on the input sketch drawn by the designer in a single angle. We thus trained the system to learn a pixel level mapping across different angles.

While the generated prototypes did have some artifacts, particularly due to lack of paired images across multiple angles, the semantic information of the shape was largely captured by the models. We are currently working on improving the accuracy of the models and generating an interactive 3D prototype of a car purely from the single angle 2D sketch input by the designer.

We developed models that visualize the car's left angular front view, left view, left angular rear view and front view. By flipping these views, we obtained right angular rear view, right view and right angular front view. The input sketch and the generated views and perspectives are shown in Fig. 12.

The rear views in particular have noticeable artifacts. However, the system was able to estimate the probable shape of the car with no information about the same in the input sketch.

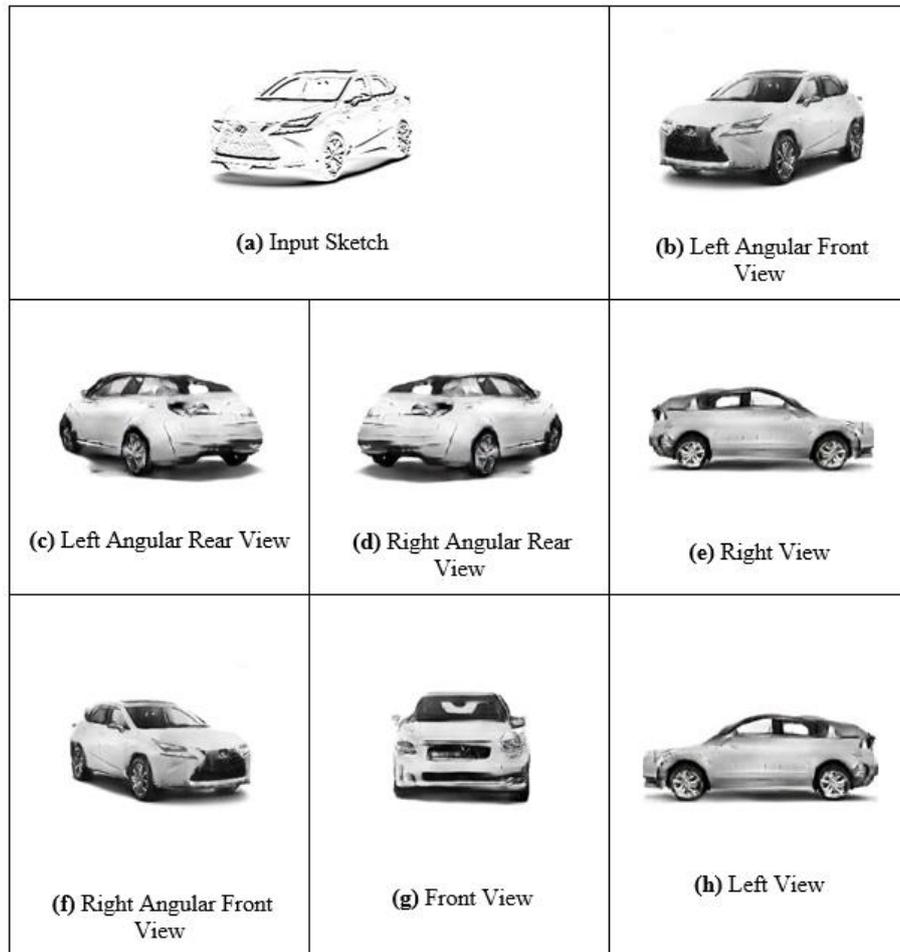


Fig. 12. Generated Unseen Car Perspective Visualizations From a Single Angle Sketch

8 Experiments on Unseen Domains

We tested our system on sketches from completely unseen domains. We first tested the system on a bike sketch (Fig. 13), as it still falls under the automobile family. The generated image suggests that the system can produce satisfactory results for other types of vehicles as well. We then tested the system on a completely unrelated domain by feeding it a sketch of a tree (Fig. 14), and the result resembles a tree in autumn. The fact that our model was not exposed to either of these domains during training phase shows that our system can easily be adapted for most industries that require creativity and design capabilities.



Fig. 13. Output of grayscale model on bike sketch



Fig. 14. Output of multi-colour model on tree sketch

9 Conclusion

In this paper, we have made three major contributions. First, a Generative Adversarial Network based system that converts car sketches to multiple prototypes having different colours and perspectives. Second, a novel convolutional sketching filter that produces sketches similar to those drawn by designers during rapid prototyping and third, a comprehensive paired dataset of about 100,000 car images (with transparent background) and their respective sketches.

We also showed certain interesting test results wherein the system was easily able to adapt to other domains that require creative intelligence and design capabilities. In most cases, human thought process is based on a recent set of prior learnings, experiences and observations. Using the proposed system to generate multiple novel combinatorial designs, the creativity of designers is augmented beyond standard creative bottlenecks.

10 Learnings and Future Enhancements

Our system qualitatively derives car designs from sketches by augmenting human creativity and intelligence using the fields of digital image processing, computer vision and (deep) machine learning. We believe, some of the possible future enhancements of our work can include:

- 3D rendering of a car from a studio sketch given a specific perspective of the car. In section 7, we showed interesting results of the system generating visual prototypes in multiple angles from a single angle input sketch. We can extend this concept by feeding the multiple generated 2D perspectives of a car to a 3D image stitching algorithm to produce a 3D rendered car from a single angle 2D car sketch.
- Producing augmented reality, virtual reality and mixed reality experiences of the generated car prototype. We can extend 3D rendering of a sketch to producing immersive experiences. Essentially, the designer will input a 2D sketch of a system and (contingent on AR/VR hardware availability) the designer can view how the generated car will look like upon production using virtual reality or other similar visualization forms.
- Generating multiple perspectives using two sketches instead of a single sketch. Currently the back side of the car is approximated as no data is available in the input sketch. To overcome this drawback, sketches of left angular front view and left angular rear view can be used together to provide almost all the required details needed to generate all perspectives of a car.
- Making the system flexible by enabling finer control of the desired generated prototype. Currently our system allows the user to choose from a certain set of colours or colour combinations. We can enhance this feature by providing the artist more granular control by allowing selection of a particular colour for a given car section. As an example, this will allow the artist to choose a different colour shade for the door and the front of the car.
- Using a crowdsourcing based approach to obtain evaluation metrics for the car prototypes obtained from the various models. Preferably, obtain metrics from a group of artists/experts and use these metrics as a means for improving certain models.

References

1. Bouchard C, Aoussat A and Duchamp R (2006) Role of sketching in conceptual design of car styling. *Journal of Design Research*, Volume 5, No. 1, pp 116–148
2. Y LeCun, L Bottou, Y Bengio, P Haffner (1998) Gradient-based learning applied to document recognition. In : *Proceedings of the IEEE*, Volume 86 Issue 11, pp 2278-2324
3. A Krizhevsky, I Sutskever, Geoffrey E Hinton (2012) Imagenet classification with deep convolutional neural networks. In : *Proceedings of the 25th International Conference on Advances in neural information processing systems*, Volume 1, pp 1097-1105

4. I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, A Courville, and Y Bengio (2014) Generative adversarial nets. In : Proceedings of the 27th International Conference on Advances in neural information processing systems, pp 2672-2680
5. Alec Radford, Luke Metz, and Soumith Chintala (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434
6. M Arjovsky, S Chintala and L Bottou (2017) Wasserstein gan. arXiv preprint arXiv:1701.07875
7. M.-Y. Liu and O Tuzel (2016) Coupled generative adversarial networks. In : Proceedings of the 29th International Conference on Advances in neural information processing systems. pp 469–477
8. P Isola, J.-Y. Zhu, T Zhou, and Alexei A. Efros (2017) Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004v2
9. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros (2018) Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593v4
10. Luke Olsen, Faramarz F. Samavati, Mario Costa Sousa, Joaquim A (2009) Sketch based modeling: A survey, Computers & Graphics, Volume 33 Issue 1, pp 85-103
11. Kang, W. C., Fang, C., Wang, Z. and J. McAuley (2017) Visually-aware fashion recommendation and design with generative image models, arXiv:1711.02231v1
12. Parui S and Mittal A (2014) Similarity-invariant sketch-based image retrieval in large databases. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV, Lecture Notes in Computer Science, Springer, Cham Vol 8694, pp 398-414
13. Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman and Alexei A. Efros (2016) Generative Visual Manipulation on the Natural Image Manifold, European Conference on Computer Vision (ECCV), arXiv:1609.03552v2
14. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei (2009) ImageNet: A large-scale hierarchical image database, Dept. of Computer Science, Princeton University, USA, Computer Vision and Pattern Recognition
15. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015) In: Deep residual learning for image recognition, Computer Vision and Pattern Recognition, arXiv:1512.03385
16. John Canny (1987) A computational approach to edge detection. In: IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. Pami-8, No. 6, pp 184–203
17. Holger Winnemoller, Jan Eric Kyprianidis, Sven C. Olsen (2012) XDoG: An eXtended difference-of-Gaussians compendium including advanced image stylization. In Computers & Graphics, Vol.36, Issue 6, pp. 720–753
18. Jin Zhou and Baoxin Li (2005) Automatic generation of pencil-sketch like drawings from personal photos. In: Multimedia and Expo, ICME, IEEE International Conference, Amsterdam, Netherlands
19. Carsten Rother, Vladimir Kolmogorov, Andrew Blake (2014) “GrabCut” — Interactive Foreground Extraction using Iterated Graph Cuts. In: SIGGRAPH, ACM Transactions on Graphics, Los Angeles, California, pp 309-314