Proceedings of the ASME 2021 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2021
August 17-20, 2021, Online, Virtual

DETC2009/MESA-12345

DESIGN FORM AND FUNCTION PREDICTION FROM A SINGLE IMAGE

Kristen M. Edwards*
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: kme@mit.edu

Vaishnavi L. Addala
Dept. of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: vaddala9@mit.edu

Faez Ahmed
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: faez@mit.edu

ABSTRACT

Estimating the form and function of a design in the early stages can be crucial for a designer for effective ideation. Humans have an innate ability to guess the size, shape, and type of a design from a single view. The brain fills in the unknowns in a fraction of a second. However, humans may struggle with estimating the performance of designs in the early stages of the design process without making prototypes or doing back-of-the-envelope calculations. In contrast, machines need information about the full 3D model of a design to understand its structure. Machines can estimate the performance using pre-defined rules, expensive numerical simulations, or machine learning models. In this paper, we show how information about the form and function of a design can be estimated from a single image using machine learning methods. Specifically, we leverage the image-to-image translation method to predict multiple projections of an image-based design. We then train deep neural network models on the predicted projections to provide estimates of design performance. We demonstrate the effectiveness of our method by predicting the aerodynamic performance from images of aircraft models. To estimate ground truth aerodynamic performance, we run CFD simulations for 4045 3D aircraft models from the ShapeNet dataset and use their lift-to-drag ratio as the performance metric. Our results show that single images do carry information for both form and function. From a single image, we are able to produce six additional images of a design in different orientations, with an average Structural Similarity Index score of 0.872. We also find image-translation methods provide a promising direction in estimating the function of design. Using multiple images of a design (gathered through image-translation) to predict design performance yields a recall value of 47%, which is 14% higher than a base guess, and 3% higher than using a single image. Our work identifies the potential and provides a framework for using a single image to predict the form and function of a design during the early-stage design process.

INTRODUCTION

Designers often begin a concept with a single sketch. From this sketch they brainstorm, add detail, and then generate a prototype. The goal of the prototype is to assess how effective a design is, often through a set of performance metrics such as weight, strength, displacement, lift, or drag. The information gathered from the prototype is critical for revising and improving a design.

We often rely on simulations to generate these metrics, because the metrics are determined by the 3D form and characteristics (such as material) of a design. Humans have an innate ability to guess the size, shape and the type of a design from a single view, but we may struggle when projecting this view into three dimensions and estimating the function or performance from said projections.

Consequently, we utilize the computer aided design (CAD) or physical prototyping process to fully assess the effectiveness of a design. However, this process is iterative, and it can be both

*Address all correspondence to this author.

time consuming and expensive, which is exacerbated through repetition. What if designers could quickly acquire the performance metrics of their design, not from a prototype, but from a single image?

FIGURE 1. Early-stage aircraft designs. A designer may struggle with estimating which design has the best and worst aerodynamic performance. We propose machine learning models, which enable performance estimation from a single image.

Consider the three images shown in Figure 1. A trained designer may be able to infer from these images alone that design a) will have better aerodynamic performance than design b), but what about design c)? During the conceptual design phase, the goal is to ideate creatively and also ensure that designs perform their desired function optimally. These goals may hinder each other if feedback on the function requires detailed, time intensive, or expensive simulations. Therefore, we propose a method that utilizes a series of machine learning models to provide designers feedback about form and functional performance in the early-stage design process, using a single image of a design.

Prior work has shown success in using generative adversarial networks (GANs) [1] to predict performance metrics like drag coefficients from 3D models [2]. These predictions are intended to classify whether a design performs its function, say flying for a plane. Feedback like this helps designers assess their design's effectiveness in early stages and make adjustments accordingly. The 3D models currently being used for these predictions are inherently more information rich than a single image. However, creating a 3D model is time consuming and requires training. In this paper, we leverage the accomplishments in the field of image-to-image translation, which have shown success in translating a single image into different orientations, thus gaining information about the form of a design [3, 4].

Our work builds from these successes. Starting with a single image of an aircraft, we perform image-to-image translation and generate images of that aircraft in six different orientations. In parallel, we use computational fluid dynamics (CFD) simulations to determine the ground truth lift-to-drag ratio for 4045 aircrafts from the ShapeNet dataset [5]. We use these ground truths to perform supervised learning of a deep neural network classification model. Our model is trained to predict the relative lift-to-drag ratio of an aircraft design from both single and multiple images of the design. We successfully use both the original input image and a combination of the original input image and the six generated images to predict the performance of an aircraft design.

Our research is some of the first work to utilize a single image for predicting the form and function of a design. Our results indicate that there is promise in this field, and we provide a framework and dataset to be used for future work. Our goal is to develop an architecture that enables human designers to ideate on the form of a design, while machine learning models provide feedback on the function.

Our contributions are listed below:

1. We demonstrate that image-to-image translation methods are capable of providing designers with additional information about the form of a design. We generate multiple distinct views of a design from a single image input. We show that the quality of predicted images can be quantified using the SSIM image similarity metric.
2. We simulate the performance, as measured by lift-to-drag ratio, of 4045 3D CAD aircraft models using a high-fidelity CFD solver. An ancillary benefit of our work is that we release a dataset with the lift and drag values of 4045 3D aircraft models calculated using OpenFOAM software for other researchers to study aircraft design.
3. We show that a single image can provide low-fidelity function feedback to aid designers in early stage conceptual design. We estimate the relative lift-to-drag ratio of an aircraft from a single input image using classification models with an average recall of 44%.
4. We demonstrate that image-to-image translation guided form prediction leads to a further 3% improvement in recall score, a final recall score of 47%, compared to prediction with a single image.

RELATED WORK

Our work touches upon three major areas of research: performance simulation in the early design stage, image-to-image translation, and classification modeling for performance estimation of designs. The great strides in these fields have motivated and provided foundations for our work. In this section we will discuss related work in each field.

Performance Simulation in the Early Design Stage

More than seventy percent of the life-cycle costs of a product are determined by decisions made by designers during the early design stages [6]. It also becomes increasingly difficult to achieve performance gains in the later stages of design. Therefore it is important to estimate and optimize performance and cost of a design as early and as accurately as possible. In this paper, we focus only on the performance estimation, however, the method we propose can generalize to other measures like cost too.

Fields, like building construction and architecture, have adopted methods for performance simulation feedback in the early design stage using computational modeling [7, 8]. These models have the benefit of providing feedback to designers before additional time and resources have been invested, or permanent design decisions have been made. Because of the benefit that early stage design performance has in the building construction field, building performance simulation (BPS) and building information modelling (BIM) methods are commonly employed [9]. However, such methods are less common in many engineering design problems, as the early design stage often focuses on expressing ideas using sketches or images.

In other engineering fields, guidelines or empirical design rules are also used to estimate performance or requirements in the early stages. For instance, when designing a truss-structure, one may use a rule-of-thumb to have more elements near the area where a force is applied, without running any simulations. Similarly, as demonstrated in Figure 1, trained designers may know that a more streamlined fuselage with wider wings will generate more lift and less resistance in the air. Therefore aircraft a) likely generates more lift than aircraft b). However, when presented with a third option, the relative comparison may become more difficult. Similarly, for decisions on novel design or decisions that are not easily visualized, designers may have to create a 3D model and simulate it using an expensive CFD simulation to figure out the relative performance of design options. In such cases, designers can benefit from simulation based performance feedback.

There has been recent success in using 3D models to predict the drag coefficient of aircrafts [2]. Additionally, convolutional neural networks (CNNs) have been successfully utilized to assess the performance of existing designs using images [10, 11]. Research has also been done to optimize the aesthetic form of a design through machine learning to enhance consumer interest [12]. However, to the best of our knowledge, no prior research has used single images of a complex 3D design to predict CFD-based performance metrics of a design in the early design stage.

Image-to-Image Translation

Image-to-image translation is a transfer learning approach in which a machine learning model learns from multiple domains simultaneously and transfers knowledge from one domain to another. The typical goal of image-to-image translation is to find a mapping of an image from one domain to another. Recent developments that use deep learning for image-to-image translation have shown excellent results. For example, for a given input image of a city in the daytime, one may use an image-to-image translation method to generate an image of what the city may look like at night time. Another example of two image domains is grayscale images and colorized images; an image-to-image translation method may be used to colorize an image from one domain, even when the method has never seen an example of the colored image of the same object in another domain. These techniques have also been used to change a certain aspect of an image. This could mean changing the hair color of a person in an image from brown to blonde [13] or changing the style of an image from photo-realistic to a Van Gogh painting [14]. Pioneering papers in this field displayed the ability to translate images from aerial views to street maps, from day images to night, and from winter scenes to summer ones [4].

Building from these initial successes, researchers have also tackled the problem of translating across different views or 3D projections, i.e. translating an image of a side-view of a car into an image of the front-view of a car [3, 13]. From a machine learning modeling perspective, Image-to-image translation can often be represented as an encoder and a decoder model. A model takes in an image from Domain $\mathcal{X}$ as the input, encodes this image into a latent representation, and then decodes this latent representation as an output image in Domain $\mathcal{Y}$. However, translating images requires a deeper understanding of the size, shape and form of an object, which makes it more difficult than other translation tasks. In order to successfully translate across different geometries, Gonzales-Garcia [3] proposed a new architecture which imposes a specific structure to the latent representation that allows different aspects of an image, like shadows, viewpoint, and orientation to be disentangled from one another. The model effectively disentangles the latent representation into three distinct parts: one part that is shared across domains and two parts that are exclusive to each domain [3]. In this paper, we use their model for image-to-image translation.

Another success within image-to-image translation is the ability to translate between sketches and images. Recent papers have shown the ability to translate from incomplete edge maps, automatically augment data, and include user interaction in generating images from sketches [4, 15, 16]. These sketch to image methods are of particular interest to our research since most early-stage design concepts are sketches which supplement our work by providing an image input. Our work focuses on predicting performance metrics from an image.

Computational Fluid Dynamics-based Performance Evaluation

Performance evaluation of complex structures (such as aircraft) using CFD simulations can be extremely time-consuming. When running CFD simulations researchers often wait for several hours and even days to obtain their results. Then, if they have to make a slight change in their input parameters, they must run their analysis again. This iterative process of defining a problem and evaluating solutions may take several days and sometimes several weeks. As an alternative to mesh-based simulations, researchers have developed various approximation-based models

that predict the results of the time-consuming analysis and reduce computation time. These simple analytical models, known as "meta- or surrogate models," are based on data available from limited analysis runs. These "models of the model" seek to approximate computation-intensive functions within a considerably shorter time than expensive simulation codes that require significant computing power. However, the surrogate models also receive the same input (typically a 3D model) as the CFD model.

It is important to note a distinction between our work and surrogate models which predict CFD performance metrics from a CAD model input. Our work deals with the problem of the input design being represented in a different form (image or sketch) than a 3D model. One can argue that an image cannot have a performance metric, as a simulation model cannot be used directly on it without a prototyping or CAD modeling step involved. While the performance of a design corresponds to its high-fidelity CAD model, we assume that the same performance is associated with the conceptual image. As we discuss later, we also look at the problem of performance prediction from a different lens: instead of training surrogate models, we train classification models that can estimate the relative performance of designs. To train such models, we first estimate 3D model performance using a CFD solver, named OpenFOAM.

OpenFOAM is a robust tool for running CFD simulations. Written in C++, it provides the framework for streamlining highly customized simulations. We modeled our simulations based on the work done by [17] in OpenFOAM. Our goal was to create a new model to learn how to predict the lift and drag coefficients. Theoretically, if a 3D model is available, one could use OpenFOAM to predict its performance metrics. We would greatly benefit from a machine learning model for the following reasons: generated 3D models might have some irregularities like holes which could vastly alter the results of a CFD simulation or prevent the simulation from executing fully; eventually our tools are meant for designers, and we do not want to place the overhead of downloading and setting up an additional application. Additionally, CFD simulations take a long time to set up and run, and we would like to provide quick feedback to designers. For these reasons, we utilized OpenFOAM to get ground truth lift and drag coefficients for the 4045 aircraft models in ShapeNet and trained a machine learning model to function in the place of OpenFOAM using this data.

METHODOLOGY

Our architecture starts with a single input image of a design, from which we generate multiple images of this design at different orientations, and finally estimate the aerodynamic performance of the design using machine learning. This process is shown in part a) of Figure 2. Each step is defined as a separate module. In this section we describe the methodology for Module 1: Image-to-Image Translation and Module 2: Machine Learning based Performance Estimation, as well as the dataset we utilized and the one we created.

Dataset

We developed our dataset using ShapeNet [5], which is an annotated, large-scale dataset of 3D shapes including aircrafts. ShapeNet provides 3D models and corresponding 2D images of thousands of aircrafts. We utilized this dataset to train our image-to-image translation model. Of the 4045 total models ShapeNet provided, which we also use for CFD simulation, 1282 aircraft models had the corresponding 2D images we required. Each of these aircraft models includes 2D images in different orientations, including various isometric views, side, top, bottom, front, and back views. We compiled the provided images for the image-to-image translation experiments shown in Figure 5.

Module 1: Image-to-Image Translation

In this module we use a single 2D image of an aircraft to generate multiple images in different orientations. Specifically, we start with an isometric image, and from that generate top, bottom, and side views, along with three different isometric views. We accomplish this using the image-to-image translation technique adopted from Gonzalez-Garcia et al. [3].

The models work by separating the latent representations of the aircrafts into the two parts that are exclusive to each domain (the image orientation), and single part that is shared by the two domains (the color and style of the image). This technique builds upon Isola et al.'s Pix2Pix framework but disentangles the style from the orientation of each image to allow for geometric changes previously not captured in image-to-image translation [4].

Our input for the image-to-image translation model is an isometric view, as indicated in Figure 5. In Figure 2 part a) we show a sketch of an aircraft connected to the aircraft image by a dashed line. This is to indicate the potential of inputting a sketch and using image-to-image translation to generate the type of input image we actually use. Our model, however, starts with an isometric 2D image from the ShapeNet dataset.

We trained six separate image-to-image translation models using Gonzalez-Garcia et al.'s framework [3] to generate six 2D images of the aircraft in different orientations given the single input image. For each model, the input image comes from the same domain, $\mathcal{X}$, however the output image domain, $\mathcal{Y}$, is different for each of the distinct models as it represents the desired output orientation. We illustrate this in Figure 5.

For each model we performed an 80-20 train-test split on a complete dataset of paired images from domains $\mathcal{X}$ and $\mathcal{Y}$. The image-to-image translation model learns to translate between the the two different orientations and generates output images in domain $\mathcal{Y}$ as seen in Figure 5.

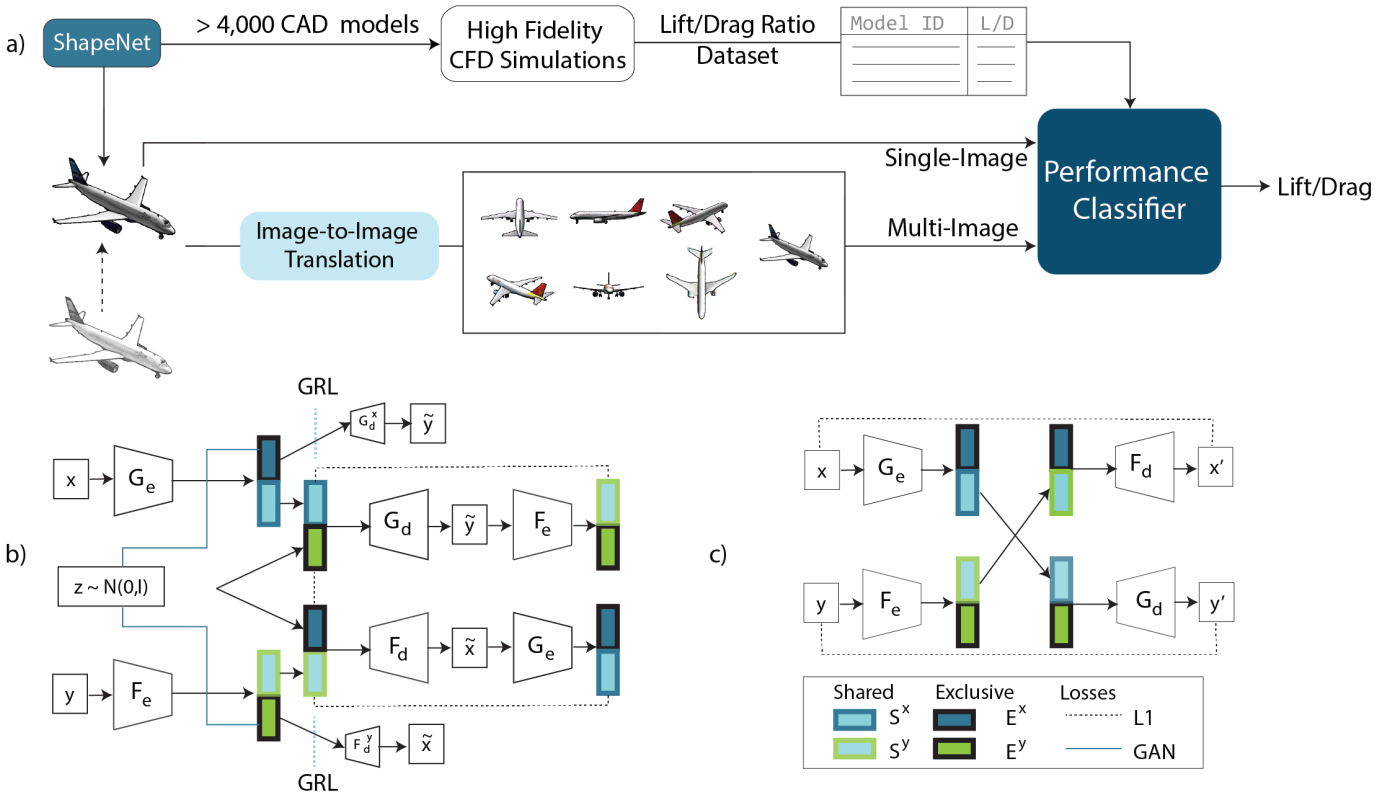

FIGURE 2. Part a) shows the overall architecture of our model. Our work is comprised of three main parts: 1) the generation of a dataset of lift and drag performance metrics, 2) an image-to-image translation model that takes in a single image of a design and produces six more images in different orientations, and 3) a performance classification model trained on the generated dataset from part 1 to predict performance metrics from one or more design images. Part b) shows the architecture of the image translators of Module 1, which successfully translate across structural changes between domains $\mathcal{X}$ and $\mathcal{Y}$. Part c) shows the architecture of the cross-domain autoencoders used in Module 1, which help ensure that the generated output still represents the information from the input image, just in a different domain.

Part b) and c) of Figure 2 show the architecture of the image-to-image translation model including both b) the image translators and c) the cross-domain autoencoders. Interested readers can explore this specific architecture further in [3]. The overall loss function to be minimized during model training is defined as:

$$\mathcal{L} = w_{\text{GAN}}\left(\mathcal{L}^{\mathcal{X}}_{\text{GAN}} + \mathcal{L}^{\mathcal{Y}}_{\text{GAN}}\right) + w_{\text{Ex}}\left(\mathcal{L}^{G^{\mathcal{X}}_d}_{\text{GAN}} + \mathcal{L}^{F^{\mathcal{Y}}_d}_{\text{GAN}}\right) + w_{\text{Ll}}\left(\mathcal{L}_S + \mathcal{L}^{\mathcal{X}}_{\text{auto}} + \mathcal{L}^{\mathcal{Y}}_{\text{auto}} + \mathcal{L}^{\mathcal{X}}_{\text{recon}} + \mathcal{L}^{\mathcal{Y}}_{\text{recon}}\right) \quad (1)$$

Module 2: Performance Estimation from a Single Image

In this module, we predict relative lift-to-drag ratio classifications of an aircraft design using a model trained on multiple views of an aircraft. We hypothesized that a machine model would be able to look at aircraft images and predict whether they have high or low lift and drag coefficients metrics, similarly to experts familiar with aircraft design.

We pose this as a classification problem and use ResNet-50 [18] to generate vector embeddings of our aircraft images. These embeddings serve as the input to our model. The ground truth is the lift-to-drag ratios calculated using OpenFOAM, run through a classification split. Similar to [17], we use the incompressible simpleFoam solver, with Reynolds-averaged simulation turbulence modeling. We run the simulation for 200 iterations and record the coefficients on the last iteration. We simulated with air velocity of 150 m/s, angle of attack of 0 degrees, kinematic viscosity of air of $\nu = 1.5e-5\ m^2/s$. The values of the lift and drag coefficients are normalized to values between 0 and 1 for before being split into categories.

Our ResNet-50 based model is a deep neural network with softmax as the final activation layer, which creates a classification model. For classification, we bin our performance values

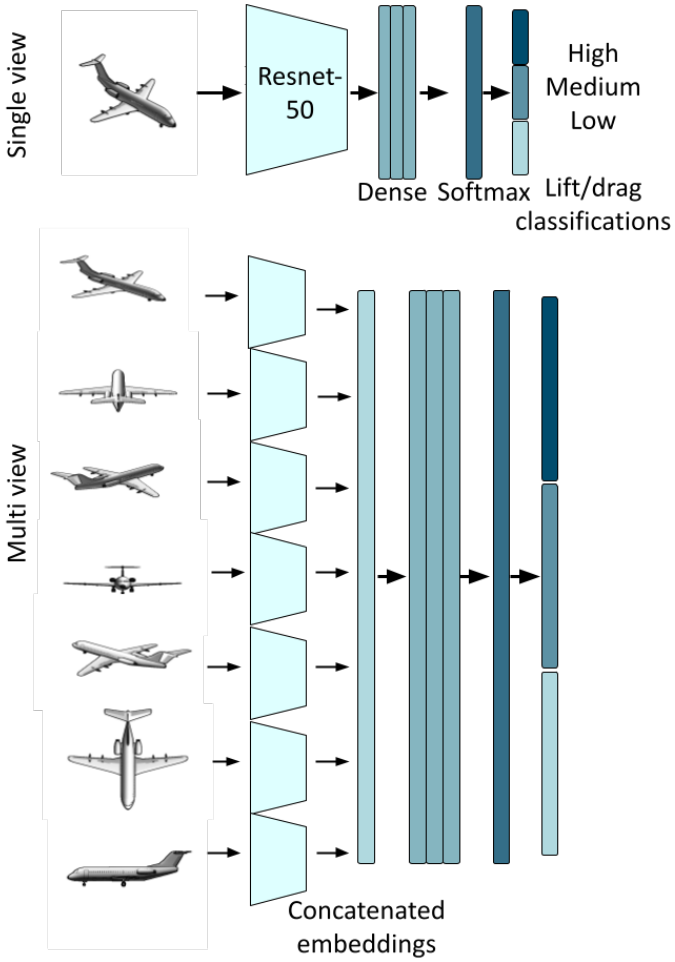

FIGURE 3. Two models with the same core architecture are shown: one for single view prediction and one for multi-view prediction. The embeddings from running aircraft images through ResNet-50 minus the last softmax layer serve as the input for the single view model. For the multi-view model, the ResNet-50 embeddings are concatenated before being passed into the dense layers.

into three groups. We classify each model by its relative lift-to-drag ratio: normalized ratios below 0.40 were classified as poor, those between 0.40 and 0.46 as medium, and those at or above 0.46 as high. These cut-offs correspond to roughly a three way split of the ratios in our data.

Our model architectures, depicted in Fig. 3 makes use of the pretrained ResNet-50 model. We run our normalized images through ResNet-50 and remove the last softmax layer to obtain embeddings of each image. For the single view model, we use a single embedding and passed it through three dense layers. For the multi-view model we concatenated the embeddings of the six total views and then ran the resulting vector through the same set of dense layers. We train the single view model with an isometric view provided in the ShapeNet dataset. This serves as the initial image. For the multi-view model, we use predicted images from module 1 for the remaining six views. We use mean squared error loss, the Adam optimizer, and a learning rate of $1e-3$. Our hypothesis was that the ResNet-50 embeddings would contain rich data about edges, their relationships, and other spatial data which would be useful for learning to predict classifications of lift to drag ratios.

Given that an image has limited information about the 3D structure of an object, we do not expect the models to provide accurate performance predictions. However, we are looking more for relative accuracy as opposed to numerical precision. The reason for this is two fold. First, for our overarching goal we mean to provide relative feedback on designs. It is important that our methodology is able to correctly predict the performance of a design relative to other designs. While very accurate lift to drag ratio prediction would provide this, we have to consider our pipeline starting from a single image that will contain propagated error. This ties to the second point of the feasibility of calculating very accurate ground truth coefficients. In ShapeNet, there are many different kinds of aircrafts. They fly at different speeds or different altitudes and have different sizes. However, it is a challenging task to automate taking all these into account while performing OpenFOAM simulations. For these reasons, we kept the simulation settings the same across all ShapeNet models and focused on relative results. One such visualization of OpenFOAM simulation is shown in Fig. 4 with the pressure distribution across the surface of an aircraft. The normal and tangent components of the pressure to the surface of the plane are integrated to produce the lift and drag coefficients, respectively.

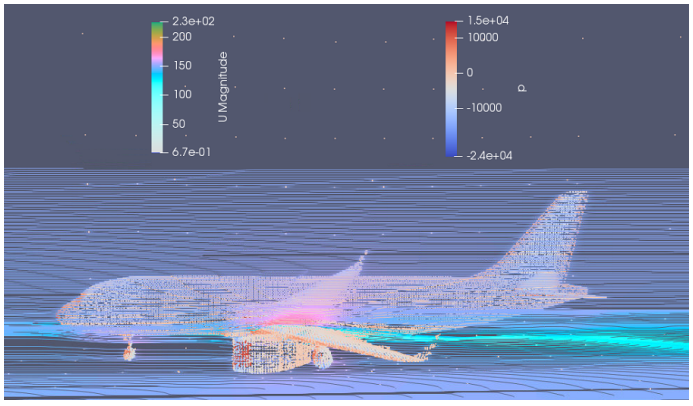

FIGURE 4. A visualization of the airflow across an aircraft during our OpenFOAM simulation. The magnitude of the velocity, U, is pictured. The pressure distribution across the aircraft model can also be seen. This is a higher resolution simulation than the simulation results used in this study. This is presented for better visualization, as apart from the granularity of meshing and scaling, the methodology is the same.

Evaluation metrics

In this section, we describe the metrics used to evaluate different models proposed in our methods.

Measuring image-to-image translation performance: As a quantitative measurement of how well our generated images matched their ground truth counterparts, we utilized the Structural Similarity Index (SSIM) [19]. A traditional objective metric of image comparison is the mean squared error (MSE). However, we chose to use SSIM rather than MSE because MSE evaluates the difference between pixel intensities in images, which does not necessarily capture structural information of an image. Our goal in Module 1's image-to-image translation is to translate across domains of different orientations, thus achieving the correct structural information of the target domain is our main objective. Wang et al. [19] developed SSIM as an objective image quality metric that is based on the similarity of structural information between two images. Therefore, we evaluate the effectiveness of Module 1 by finding the SSIM score between our generated images and the ground truth images. SSIM scores range from 0-1, with 1 indicating perfect structural similarity between two images.

Measuring classification performance: In Module 2 we both generate a dataset of performance metrics (lift, drag, and lift-to-drag ratio) for the 4045 aircraft models in ShapeNet, and build a supervised deep neural network classification model that predicts the relative lift-to-drag ratio of an aircraft.

We alter the traditional evaluation metrics of binary classification, precision and recall, for our multi-class problem. Precision and recall are both defined based on the number of true positive (TP), false positive (FP), and false negative (FN) prediction. For binary classification between classes -1 and 1, TP means predicting 1 and the actual class being 1. FP means predicting 1 but the actual class is -1. FN means predicting -1 when the actual class is 1.

Precision indicates how many positive predictions are true. It is defined as:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall, also known as the true positive rate (TPR), measures how many of the positive cases our model is able to correctly predict. Recall is defined as:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

To use these metrics for multi-class classification, we simply take a one versus all approach. For each class we find the precision and recall by assuming that class is positive and all other classes are negative. We also calculate the average recall for all classes.

RESULTS

In this section we detail the results of both modules using qualitative examples and quantitative metrics described in the Evaluation Metrics section above. Due to the sequential nature of our model, the results of Module 2 rely on the results of Module 1; thus, both accuracy and errors propagate throughout the modules.

Image-to-Image Translation Yields Additional Form Information

Results for the image-to-image translation from one orientation to another are shown in Figure 5. This figure shows the qualitative results of Module 1. The input image shown in part a) of Figure 5 is an isometric view of an aircraft. From this input image, six different image-to-image translation models were trained to translate the input into six different orientation domains, each shown in the outputs column.

Part b) of Figure 5 demonstrates examples of generated output images, their ground truth counterparts, and the SSIM score that pair received. The SSIM score, as described in the Evaluation Metrics section above, is a quantitative measure of the structural similarity of two images. We show three examples of different pairs and their SSIM score. The examples shown have decreasing SSIM scores from the top row to the bottom. The first row shows a generated image and ground truth pair with the highest SSIM score of 0.979, the second row has an SSIM score or 0.872, and the third row has the lowest SSIM score of 0.803.

The SSIM scores of the first two rows are quite high, indicating high structural similarity between the generated images and their ground truth, which can be visually verified too. The low score of the third row demonstrates a common theme in Module 1. The plane model used in the third row is less like the classic aircraft shape. In general, the image-to-image translation performed the best on aircraft models that looked like the models in the first two rows. For models that differed greatly from this design, the image-to-image translation was often unable to produce realistic results. For example, the image in the Generated column of row three has a part of the plane disconnected from the rest.

SSIM Scores Vary Based on Orientation

The effectiveness of Module 1 is further reiterated in Figure 6. Figure 6 shows a violin plot of the SSIM scores for the 252 generated images in each of the six translated orientations. The overall average SSIM score for all 1512 generated images is 0.872. To gain a visual understanding of what this score might

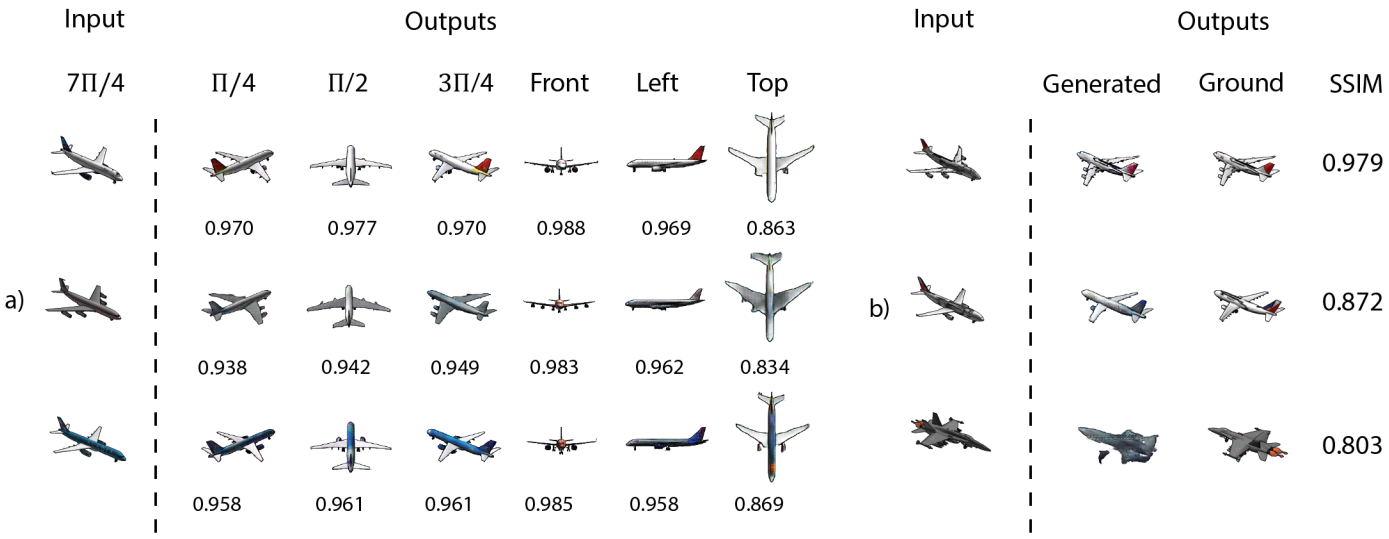

FIGURE 5. Results of the image-to-image translation. Part a) shows the results of image-to-image translation on dataset one, the 2D images provided from ShapeNet. The input is an isometric image of an aircraft, and the outputs are generated images of that same aircraft model in six different orientations: three isometric orientations, front, side, and top orientation. Under each image is the SSIM score, where a score of 1 indicates perfect structural similarity between a generated image and its ground truth counterpart. Part b) exhibits three examples of image-to-image translation from one isometric view to another. The input image is shown in the left column, the middle column shows the image generated through image-to-image translation as well as its ground truth counterpart, and the right most column shows the SSIM score of the generated and ground truth pair.

mean, one can check the middle row of Figure 5 part b), which has an SSIM score of 0.872.

Figure 6 exhibits the SSIM distribution as well as the average SSIM score for each of the predicted orientations. The results indicate that the effectiveness of image-to-image translation from one orientation (Domain $\mathcal{X}$) to another (Domain $\mathcal{Y}$) varies based on Domain $\mathcal{Y}$. For example, when predicting images in the front orientation, image-to-image translation performs consistently well, with an average SSIM score of 0.943. In contrast, our image-to-image translation produces the worst results when predicting into the top orientation. These predicted set of images can provide insights on the form of a design to a user. For each design, we now have six additional images in various orientations. In the next section, we show how these images can also help in predicting the function.

Generation of a Performance Metric Dataset

We used OpenFOAM CFD simulations to generate the lift and drag coefficients of all 4045 aircraft models in the ShapeNet dataset, and built a new dataset of the performance metrics for each model. Figure 7 presents the histograms for lift, drag coefficients from the OpenFOAM simulations. We also show the distribution of lift-to-drag ratio. The mean and standard deviation for the lift coefficients are 0.259 and 0.879. For the drag coefficients, they are 0.563 and 0.458.

Figure 8 illustrates examples of performance metric results generated through OpenFOAM CFD simulations. Each row shows three aircrafts that increase in the specified performance metric from left to right. These examples give qualitative indications of the results generated by OpenFOAM. For example, the top row looks at the lift coefficient of an aircraft, and we observe that the "low" aircraft has a less streamlined fuselage and wings than the "high" aircraft.

Machine Learning Models Yield to Function Feedback

In this section, we show our results for predicting performance evaluation starting with a single image. We used our generated dataset of lift-to-drag ratios as the ground truth to compare against our predicted lift-to-drag ratio found through a neural network.

Due to the continuous nature of the lift-to-drag ratios, we first experimented with regression models. When training our models for regression, we noticed that the models predicted the mean value of lift-to-drag ratio with very slight variation. This is a common issue in regression training since predicting the mean value corresponds to a local minimum in the loss manifold. As a result, our R^2 values were poor for both single view and multi view models - both around -0.09 for the end to end models. Due to the poor regression results and the fact that precise numbers are not our goal, but rather low-fidelity feedback for early stage design decisions, we opted for a classification model for our lift to drag ratios.

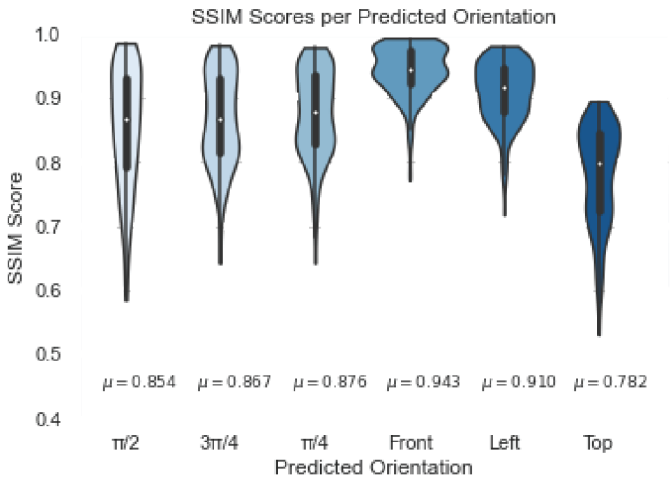

FIGURE 6. A violin plot of the SSIM scores for all six generated orientation. The input image is an isometric view, as indicated in the input column of Figure 5. For each of the predicted orientations, 252 test images were generated and their SSIM scores were found in comparison to their ground truth counterparts.The median for each orientation is shown with a white dot, the interquartile range is shown as a black bar, and each orientation's mean is shown above its name on the x-axis. The average of all SSIM scores for all orientations is 0.872, with a standard deviation of 0.085.

Our prediction is multi-class; with the classes being low, medium, or high lift-to-drag ratio. We illustrate our model accuracy with a confusion matrix, as shown in Table 1. The predicted class is along the y-axis, the actual class is along the x-axis. A perfect confusion matrix only has numbers along the diagonal, meaning every aircraft that is predicted to be in the low class, is actually in the low class – the same applies for the medium and high classes. We show both the results of predicting the lift-to-drag ratio using a) a single image and b) multiple images of a design.

In both of these cases the initial input into our model is a single image of a design. In case a) the image is directly input into our performance classifier model, as seen in Figure 2. In case b), however, the single input image is first input into our image-to-image translation model, which produces six additional images of the aircraft in six new orientations. In combination with the original input image, we input these seven images into the performance classifier to predict the lift-to-drag ratio class of the aircraft.

Our hypothesis and motivation for case b) is that providing additional information about the form of an aircraft, including front, side, and top views, will improve a model's ability to predict the aircraft's function in the form of lift-to-drag ratio. We show our results in Table 1.

The average precision with a single image is 44%, while the average precision with multiple images in 48%, producing an increase of 4% from the single image. Similarly, the average recall with a single image is 44%, while the average precision with multiple images in 47%, producing an increase of 3% from the single image.

Overall, the model with multiple views had a recall of 47% while the model with a single view had one of 44%. A random guess would give a recall of 33%. This shows that image-to-image translation is able to improve classification performance. as measured by recall, by 14%.

We note that with testing on translated views generated from the training portion of module 1 (and not on predicted translated images from the test set), the recall was 49%, 5% higher than predicting with a single image and 2% higher than when testing on the test images from module 1, which is what we are reporting. This 2% drop in performance can be attributed to error propagation that occurs through the image-to-image translation.

DISCUSSION

In Module 1 we demonstrate the ability to generate six design views in different orientations from a single image. We evaluated the effectiveness of our image-to-image translation model using an SSIM score. As demonstrated in Figure 6, the average SSIM score for all orientations is 0.872, which can be qualitatively understood by looking at Figure 5 part b) in which the middle row shows the generated image and ground truth comparison that produce an SSIM score of 0.872.

We were successful in translating a single isometric view into various views that contain dissimilar information: such as the top and front views. The ability to generate these dissimilar views is particularly important in performance prediction.

Drag generally increases with an increased cross-sectional area, which is best defined by the front view. Lift, on the other hand, may be most affected by information portrayed in the top view. Our success in generating these views indicates that the inclusion of an image-to-image translation model should increase the accuracy of performance metric predictions.

In image-to-image translation, we notice that certain orientations are most effectively predicted. For example, Figure 6 shows that the front view has the highest average SSIM score as well as the most condensed SSIM scores across all tested aircrafts. In contrast, the top view has the lowest average SSIM score. Because drag is most impacted by the front view, and lift most impacted by the top view, these translation results may create a disparity in predicting lift vs. drag coefficients. However, in Module 2 we predict the lift-to-drag ratio, which incorporates both coefficients. Future work may explore analyses of predicting these coefficients separately.

In Module 2 we demonstrate performance estimation starting with a single image. In particular, we compare the results of

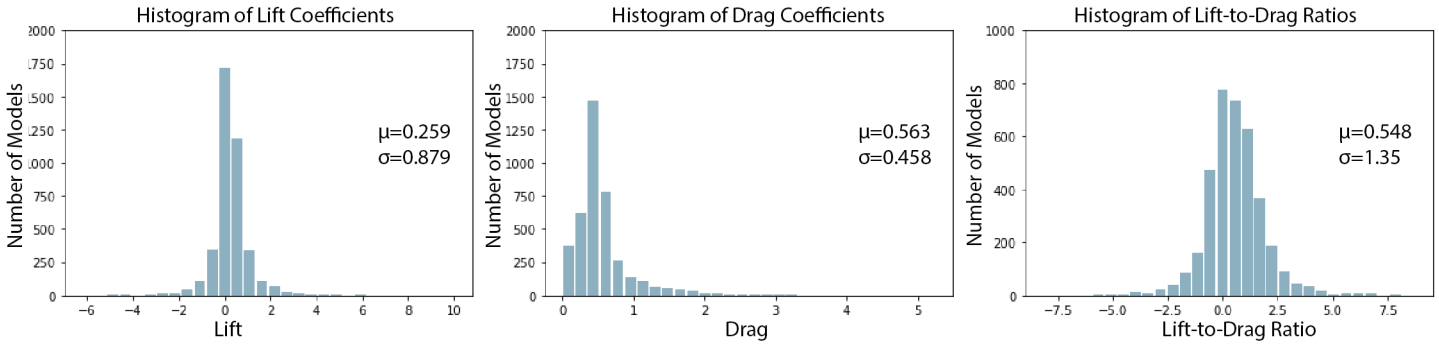

FIGURE 7. The histograms of lift coefficients, drag coefficients, and lift-to-drag ratio from the OpenFOAM simulations run on all 4045 ShapeNet aircraft models.

	Single-Image Input			Multi-Image Input		
	Low	Medium	High	Low	Medium	High
Low	48	43	37	39	25	18
Medium	29	53	38	24	33	33
High	37	34	72	6	19	36
	Actual Class			Actual Class		

Predicted Class

	Single-Image Input		Multi-Image Input	
	Precision %	Recall %	Precision %	Recall %
Low	38	42	48	57
Medium	44	41	37	43
High	50	49	59	41
Average	44	44	48	47

TABLE 1. Confusion matrix for multi-class classification. Average recall with a single image is 44%, which is significantly better than a random baseline value of 33%. We observe that multiple images further improve the classification performance with average recall increasing to 47%. Interestingly, the multi-view model performed better against extreme misclassifications. The percentage of samples misclassified from low to high and vice versa decreases when using multi-view models.

estimating the performance with a single image alone vs. with multiple images generated from the single image through Module 1. The confusion matrix in Table 1 shows our results. The average TPR for a single-image input is 44%, while the average TPR for a multi-image input is 47%. The average precision also increased from 44% to 48% with single vs. multi-image inputs. These results indicate slight classification improvements when using multiple images.

Our classes are not independent of one another, meaning that aircrafts in the low class have design metrics closer to that of the medium class than the high class. Because of this, not all misclassifications are equal. An extreme misclassification is classifying a low as a high or vice versa, we aim to minimize these misclassifications over all others. The multi-view model performed better against extreme misclassifications, only falsely predicting 2.6% of low class aircrafts as high class and 7.7% of high class aircrafts as low class. These false predictions are higher, both at 9.5%, with the single view model.

We want to highlight that our research demonstrates preliminary work in the area of predicting the form and function of a design from a single image. Our performance evaluation provides a low-fidelity prediction of the lift-to-drag ratio of an aircraft. This prediction is not our main contribution; rather, we have identified a problem and created a framework to solve it. We have shown the potential of using a single image to predict design performance, and provided preliminary results that suggest there is promise in this area. Further, we have provided a dataset for researchers to use for future work in this field.

LIMITATIONS AND FUTURE WORK

This work demonstrated that data-driven methods show promise in predicting design form and function. Information on performance can be critical for a designer to make informed changes to a design and observe how the performance may change. Similarly, richer information on the form can help them visualize different aspects of the design, without spending sig-

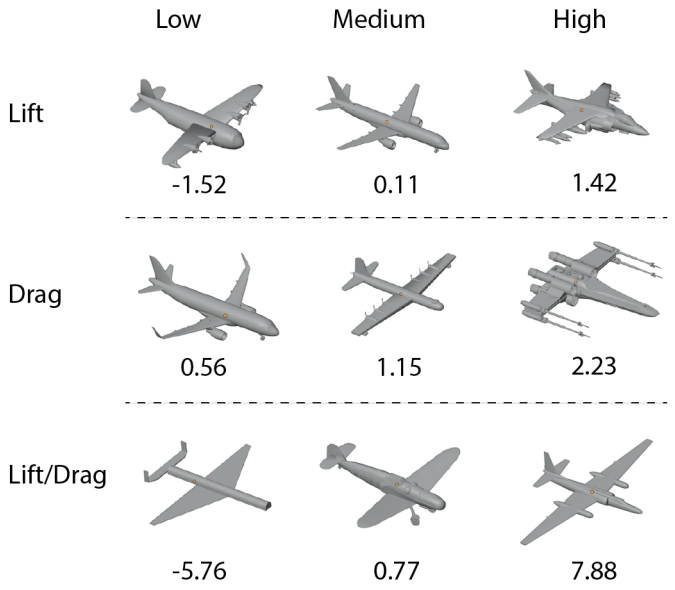

FIGURE 8. An illustration of the performance metric results we gathered through OpenFOAM CFD simulations. For each aircraft model in the ShapeNet dataset we found the lift, drag, and lift to drag ratio. Here we demonstrate example planes for different values of these metrics.

nificant time in building an accurate 3D CAD model or running complex numerical simulations. However, it is important to understand that predicting form and function from a single image should be limited to providing rough estimates and cannot replace high-fidelity modeling and simulations. The performance of the methods is dependent on both how rich the training data is and how dependent the performance is on the external form of the design, which can be observed in an image. Below, we note the assumptions and limitations of our proposed methodology.

We tried an end-to-end model for predicting lift to drag ratios, with both a single view and multiple views. These models were composed of convolutional layers, which we hypothesized would provide information on the spatial relationships between parts of the aircrafts, and dense layers. With this structure, we were unable to learn to predict lift to drag ratios from the aircraft views - it largely predicted values very close to the mean of the ratios. Our thoughts on the reasons for this are guided by the comparison of the performance of the Resnet-based model. Our dataset size for training was only around 1,000 models since those were the ones that came with standardized screenshots in ShapeNet. This is quite a small number to learn a complex physical phenomenon like lift to drag ratio. Coupled with the small dataset, variety within the dataset likely also played a role. Many aircrafts have many similar structures, making it difficult for the model to learn generalizable methods of predicting the lift to drag ratio. Due to the complexity of the problem and the small amount of data available from which to generalize, we believe reasonably training the millions of parameters was unattainable. Since Resnet was trained using a variety of different images in addition to a deep framework, the embeddings generated carried more of the pertinent information needed to calculate the ratios. Lastly, we will discuss some of the attempts we made to combat the difficulty presented in learning for our end-to-end model. We used data augmentation to randomly scale and shift aircraft views while keeping the correct ratio truth value. We thought the slight perturbations would push for more generalizable learning. This allowed us to increase our dataset size, but the learning did not significantly improve. We also tried different normalization techniques for both the images and the coefficients and/or ratios themselves. We believe that since humans can generally predict lift and drag coefficients from looking at images of aircrafts (for example, the top view is helpful for predicting drag and the front for lift), that with the right model architecture and data normalization, a neural network should be able to as well. We plan to look further into advanced techniques and dataset augmentation, to reach the limits of predicting performance from a single image.

Dependence on the number of views: In our experiments, we train our machine learning model to generate six views for each input image. As shown in Figure 6, image-to-image models vary in performance for different views. For instance, the top view was the most difficult to generate, while the front view generation quality was more easily generated. As these views are used as an input to the classification model, both the type of views (top, front, side) and their generation quality is expected to directly impact the prediction performance. In future work, we aim to conduct a systematic view to uncover design performance's dependence on different views.

Dependence on the design space: Our image-to-image translation model learns from a large collection of designs to generate different views of an image. While the overall performance of the model is good (as measured by average SSIM scores), on digging deeper into the results, we found that the performance of the image-to-image translation model drops significantly for novel designs where the look of the aircraft differs from most other aircraft. This is demonstrated in the bottom right image translation output in Fig. 5, which shows a lower SSIM score for a unique aircraft design. While the limitation of a machine learning model not performing well for sparse regions of data is widely known, this issue also limits the applicability of our model for novel input images - where knowing performance estimates can be even more important for a designer. Hence, the generalizability of our approach is limited by the design space covered by the training data. Our approach will generalize better for applications with large datasets and multiple classes of designs.

Future work: While recent advances in machine learning methods have shown the promise of performance prediction with a 3D CAD model as input, predicting complex CFD-based performance metrics from a single image appears improbable. However, our work shows that by leveraging transfer learning, machine learning models can provide good estimates of performance. Future work will focus on refining these models, training regression models, and testing the limits of their accuracy and generalizability. We will also expand our analysis to directly predict performance from human-made design sketches, instead of projection images. While a sketch may have even less information than an image, it is also closer to what a designer would normally draw in the conceptual design stage. To test our model, we will first create a dataset of sketches along with CAD models and then leverage both sketch understanding and image-translation models for performance prediction. Finally, an important area of research is to conduct experiments with humans and AI working together to test the final efficacy of providing enhanced form and function information on creative outcomes. We envision an AI-assistant, which can also recommend design changes to a human-generated conceptual design, such that the final high-fidelity model has high performance.

CONCLUSION

Our goal is to help designers generate new ideas by giving them feedback on function and form of initial designs. We have shown that employing state of the art image-to-image translation techniques during early-stage design can provide richer information about the form of a design. This is demonstrated through Module 1, in which we generate six images of a design in six different orientations from a single isometric image.

In order to provide a designer with feedback regarding the function of a design, we have 1) developed a dataset of aerodynamic performance metrics for 4045 aircrafts and 2) built a deep neural network model that provides low fidelity performance metrics of a design from a single or multiple images.

This is a preliminary work in which we believe our greatest contributions are identifying and providing methods for ongoing research. We have identified a problem- using a single image to predict the form and function of a design- and created a framework to solve it. Our results have shown the promise of using a single image to acquire inexpensive low-fidelity performance predictions in the early stage of a design. We intend for this method to provide feedback to designers without having to create complex 3D CAD models or run time-consuming CFD simulations for every small change during conceptual design stage.

We hope that our work will both give rise to and support future research in this field. We believe that using machine learning for these performance predictions can enable better human-AI collaboration. This collaboration can capitalize fully on humans' ability to extrapolate, understand, and create new forms when provided little information, and a machine's ability to rapidly evaluate function when provided more information. Utilizing these complementary abilities can enable humans to ideate effectively on the form, while AI gives feedback on the function.

ACKNOWLEDGEMENTS

We thank the Ida M. Green Fellowship for supporting Kristen Edwards's research. We thank MIT's Undergraduate Research Opportunities Program for supporting Vaishnavi Addala's research. We also thank the other members of the DeCoDE Lab for helping at various stages of the project.

REFERENCES

[1] Radford, A., Metz, L., and Chintala, S., 2016. Unsupervised representation learning with deep convolutional generative adversarial networks.

[2] Shu, D., Cunningham, J., Stump, G., Miller, S., Yukish, M., Simpson, T., and Tucker, C., 2019. "3d design using generative adversarial networks and physics-based validation". *Journal of Mechanical Design,* ***142***, 11, pp. 1–51.

[3] Gonzalez-Garcia, A., van de Weijer, J., and Bengio, Y., 2018. "Image-to-image translation for cross-domain disentanglement".

[4] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., 2016. "Image-to-image translation with conditional adversarial networks". *arxiv*.

[5] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., and Yu, F., 2015. ShapeNet: An Information-Rich 3D Model Repository. Tech. Rep. arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago.

[6] Mileham, A., Currie, G., Miles, A., and Bradford, D., 1993. "A parametric approach to cost estimating at the conceptual stage of design". *Journal of engineering design,* ***4***(2), pp. 117–125.

[7] Picco, M., Lollini, R., and Marengo, M., 2014. "Towards energy performance evaluation in early stage building design: A simplification methodology for commercial building models". *Energy and Buildings,* ***76***, pp. 497–505.

[8] Basack, S., and Sarkar, A., 2019. "Bim framework for operational energy assessment in composite climate at early design stage". *Building Engineer,* ***94***, 03, pp. 22–27.

[9] Singh, M. M., Singaravel, S., Klein, R., and Geyer, P., 2020. "Quick energy prediction and comparison of options at the early design stage". *Advanced Engineering Informatics,* ***46***, p. 101185.

[10] Shi, J., Tao, Y., Guo, W., and Zheng, J., 2020. "Cnn based

defect recognition model for phased array ultrasonic testing images of electrofusion joints".
[11] Zhang, Q., Zhang, M., Gamanayake, C., Yuen, C., Geng, Z., Jayasekaraand, H., Zhang, X., wei Woo, C., Low, J., and Liu, X., 2020. Deep learning based defect detection for solder joints on industrial x-ray circuit board images.
[12] Kang, N., Ren, Y., Feinberg, F., and Papalambros, P., 2019. Form + function: Optimizing aesthetic product design via adaptive, geometrized preference elicitation.
[13] Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J., 2017. Learning to discover cross-domain relations with generative adversarial networks.
[14] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A., 2020. Unpaired image-to-image translation using cycle-consistent adversarial networks.
[15] Chen, W., and Hays, J., 2018. Sketchygan: Towards diverse and realistic sketch to image synthesis.
[16] Ghosh, A., Zhang, R., Dokania, P. K., Wang, O., Efros, A. A., Torr, P. H. S., and Shechtman, E., 2019. "Interactive sketch & fill: Multiclass sketch-to-image translation". In Proceedings of the IEEE international conference on computer vision.
[17] Cunningham, J., Simpson, T., and Tucker, C. S., 2019. "An investigation of surrogate models for efficient performance-based decoding of 3d point clouds". *Journal of Mechanical Design,* ***141***.
[18] He, K., Zhang, X., Ren, S., and Sun, J., 2016. "Deep residual learning for image recognition". In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
[19] Zhou Wang, Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., 2004. "Image quality assessment: from error visibility to structural similarity". *IEEE Transactions on Image Processing,* ***13***(4), pp. 600–612.