

In-Context Learning and Prompt Design for Computer-Aided Aerodynamic Engineering Application

Xingyue Zhang
Supervisor: Pascal Fua
Assistant: Zhen Wei



School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

January 4, 2024

In-Context Learning and Prompt Design for Computer-Aided Aerodynamic Engineering Application

Xingyue Zhang

SCIPER: 351693

xingyue.zhang@epfl.ch

École Polytechnique Fédérale de Lausanne

Abstract

This study explores the innovative application of visual prompting via image inpainting in the field of aerodynamic shape optimization, particularly for predicting pressure fields around airfoils. Our approach utilizes a 2×2 grid structure comprising airfoil images and their corresponding pressure fields. This grid structure is then fed into a pre-trained universal visual model that can predict the pressure field of an unseen airfoil based on the provided example, without any task-specific fine-tuning or model modification. We conducted extensive experiments to understand the effects of various factors such as airfoil zoom-in level, color schemes of the pressure field, aggregation, and the similarity of airfoils used in the visual prompts. The results indicate that larger airfoils and specific color schemes enhance the model's accuracy. Aggregation also helps improve model performance. Moreover, using similar airfoils as examples significantly improves the performance compared to dissimilar ones. The code and data are available at: https://github.com/bREAKtHEDOLL/visual_prompt_airfoil.git.

1 Introduction

The development of optimum aircraft in the field of aeronautics is one of the most intricate challenges in engineering. To effectively address the various contributing disciplines, designers have to depend on automated optimization techniques.

For a long time, the industry practice of addressing design problems in aeronautics is by using a two-level approach (Giannakoglou, 2002). The initial stage involves crafting the overall system design, drawing on insights from past projects and basic modeling tools. The subsequent stage is the detailed component design, where tools as advanced as three-dimensional Navier-Stokes solvers may be employed (Giannakoglou, 2002).

Shape optimization is indeed a common challenge in aeronautics. *Computation Fluid Dynamics* (CFD) is crucial in aerodynamic shape optimization, allowing for the assessment of complex shapes' aerodynamic performance without the need for expensive experimental procedures. Integrating CFD into the conceptual design process can significantly influence aircraft design. During this early stage, critical configuration decisions are made, and global optimization becomes particularly important, guiding the design towards optimal aerodynamic efficiency.

Following the initial conceptual stage, the design process progressively shifts towards local optimization for refining the design. Traditionally, optimization in the conceptual phase primarily relies on empirical design tools, which draw from past aircraft models and the experience of engineers (Forrester et al., 2006). While discussing traditional optimization algorithms, it is important to note that their primary aim is to iteratively enhance a single individual, guided by deterministic rules. This approach means that the final outcomes are inherently dependent on the initial solution, and there is a constant risk of becoming trapped in local optima. Therefore, a significant challenge lies in efficiently executing global optimization through physics-based simulations. This efficiency is vital to enable the utilization of these methods within the constrained time frames typical of the conceptual design stage.

The application of surrogate models in optimization plays a pivotal role. These models are commonly known as approximation, meta, or response surface models. While a global surrogate model, such as one estimating an aircraft's drag, may not match the accuracy of individual simulations, it provides near-instant results (Forrester et al., 2006). This speed facilitates comprehensive global searches, making it a valuable tool in the optimization process. In other words, the surrogate model functions not as an optimizer itself, but as a means to accelerate the optimization process. However, surrogate models often require extensive data for training and usually exhibit limited generalization capabilities. Is

there a way to quickly and accurately predict fluid fields without training?

A recent paper proposed the idea of adapting a pre-trained visual model to novel downstream tasks without task-specific fine-tuning or any model modification (Bar et al., 2022). The authors took inspiration from prompting in NLP and examined *visual prompting*: given input-output image example(s) and a new query image, the objective is to generate the output image based on the given example. The authors showed that this is a problem of *image inpainting*, i.e., filling in a hole in a concatenated visual prompt image. The authors trained masked auto-encoders on a specially compiled dataset of unlabeled images taken from academic papers on Arxiv. Applying visual prompting to these pre-trained models, the authors produced impressive results for a variety of downstream image-to-image tasks, such as foreground segmentation, single object detection, colorization, edge detection, and more (see Figure 1) (Bar et al., 2022). Their approach has remarkable generalization performance without having to specify tasks manually.

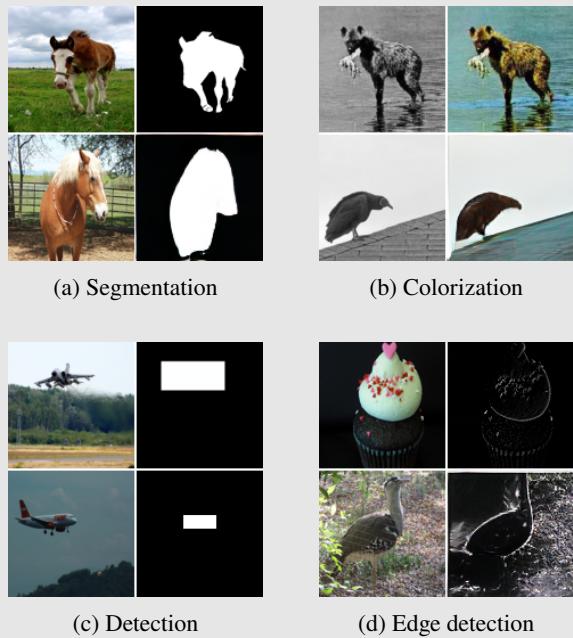


Figure 1: Visual prompting prediction examples

Let’s take a closer look at this work. Recently, self-supervised learning has become popular in computer vision and natural language processing. As modern deep learning models grow in capacity, they tend to overfit when trained on small labeled datasets. Self-supervised learning offers a remedy for this by generating “free labels” for any dataset, eliminating manual labeling and satisfying the large data needs of these models. However,

the features learned through this method usually require fine-tuning with labeled data for specific tasks. The paper by Bar et al. (2022) makes an endeavor to address this problem.

Bar et al. (2022) were inspired by *prompting* in NLP, which is to use a model for a new task without any training. For instance, the input prompt is a French-English sentence pair and a query which is a French sentence. The output would be a translated English sentence of the query French sentence. The authors applied this ideology to the visual domain and generated large-capacity image inpainting models (trained on the right data) that can perform various tasks without any fine-tuning. They constructed a grid-like single image, namely *visual prompt*, where the upper-left image is the task input example, the upper-right is the task output example, and the lower-left is the query image. The goal of the inpainting model is to predict the masked region in Figure 2 so that the lower two images are consistent with the top two images. They introduced a new dataset that enables a model to learn this grid-structure autonomously, without the need for labels, description of the task, or extra details about the grid’s configuration (Bar et al., 2022). This approach is also a form of *in-context learning*, a ‘hidden’ capability originally discovered in large autoregressive language models (Radford et al., 2021; Zhang et al., 2023).

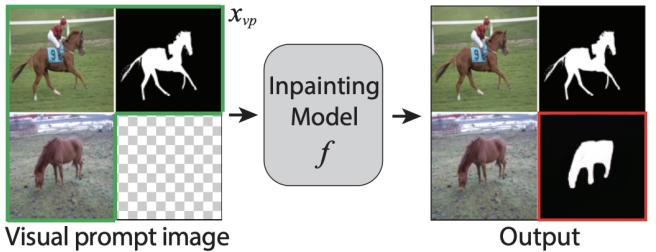


Figure 2: How the inpainting model works (Bar et al., 2022)

Bar et al. (2022) proposed an inpainting model which is a combination of MAE (He et al., 2022) and VQ-GAN (Esser et al., 2021) to implement an inpainting function f with a neural network.

In this work, we mainly focus on predicting the fluid fields around airfoils, which is essential for aerodynamic shape optimization. Inspired by the work by Bar et al. (2022), can we apply visual prompting via image inpainting to predicting fluid fields around airfoils? Traditionally, predicting fluid fields around airfoils has been a time-consuming and costly process, heavily reliant on numerous physical experiments that could span many years. Instead, can we combine Computational Fluid

Dynamics (CFD) and visual prompting (deep learning) to predict the physical measurements around an airfoil given a task example? The primary objective of this task is to predict fluid fields around airfoils quickly and reduce the number of simulations required. By enhancing the efficiency of Computational Fluid Dynamics applications, we aim to lower costs and expedite the analysis process. Through quicker and more efficient predictions, engineers and researchers can achieve more accurate designs and analyses with fewer resources, resulting in more cost-effective and time-efficient operations.

To be more specific, the 2×2 grid consists of four images:

- **Upper left:** Task input example, which is an image of airfoil A.
- **Upper right:** Task output example, which is an image of airfoil A and its surrounding pressure field.
- **Lower left:** Query image, featuring airfoil B. These first three images comprise the visual prompt image.
- **Lower right:** Model output result, which is a prediction of the surrounding pressure field of airfoil B.

2 Methods

2.1 Airfoil Datasets

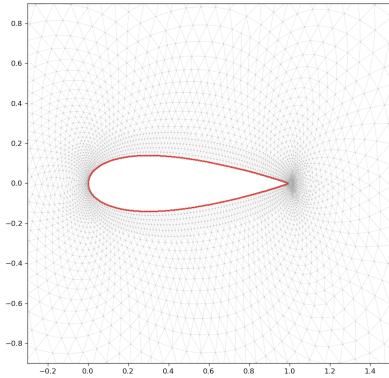


Figure 3: Visualization of data for one airfoil

Our airfoil data is sourced from *UIUC Airfoil Data Site* (Selig, 1996). We are working with two datasets. In the first dataset, each file corresponds to one unique airfoil. Figure 3 is a visualization of one such file. Each file contains a set of points that are inside, on, or outside the airfoil. We know the coordinates of each point and each point has its own ID. Moreover, notice that outside the airfoil, the points form many triangles. Three points compose three vertices of a triangle (face). To represent

such a triangle, we use the three point IDs that form the triangle and we give each triangle its own ID number.

The red contour in Figure 3 illustrates the airfoil surface, which consists of many line segments. Each line segment is defined by a pair of points, allowing us to present each segment with two point IDs. This graph additionally features a boundary, which is not depicted in the graph. The boundary is also composed of many line segments defined by pairs of points.

In the second dataset series, each file corresponds to the same airfoils as in the first set. From this series, for each airfoil, information including density, momentum, energy, pressure, pressure coefficient, temperature, and Mach values is available for every point we defined in the first dataset. We mainly use pressure coefficient values in this work.

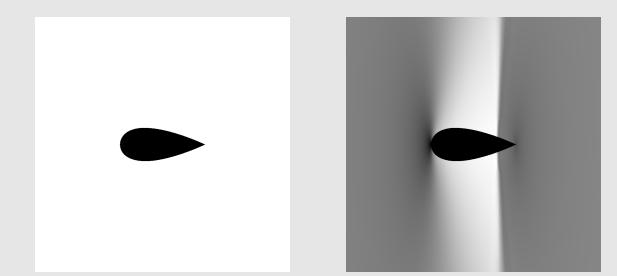
2.2 Visualizing Pressure Field Around Airfoil Surfaces

To create airfoil images that can function as visual prompt examples, it is essential to illustrate task input example, which is the airfoil shape, and task output example, which shows the airfoil shape and the pressure coefficient values surrounding it. The intensity of the pressure coefficients is represented through the varying color brightness, providing a visual indication of their magnitude.

From the datasets, we only know the pressure coefficient values for points on the vertices of triangles. How do we achieve a smooth visual representation of these values? We use interpolation to solve this issue. Given a coordinates of a point, if it lies outside the airfoil but within the boundary, we first determine the specific triangle it belongs to. Then, we calculate its *barycentric coordinates* relative to this triangle. Using these coordinates, we interpolate the pressure coefficient for this new point. This process enables us to accurately estimate pressure coefficients for areas between known vertices.

Next, we move on to the task of creating the image of an airfoil and its surrounding pressure field. An image is determined by pixels. For instance, if we define an image with a resolution of 100×100 pixels, we can then apply the above described method to derive the pressure coefficients for these points. Note that we don't need to illustrate this within the contour of the airfoil. We can select from various colors to illustrate the pressure coefficients outside of an airfoil. We can also change the color and size of the airfoil itself. The implementation details can be found in the supplementary material. The resulting visualization is shown in Figure 4b.

Now that we know how to plot the task output example, we also need to plot the task input example, which is



(a) Task input example

(b) Task output example

Figure 4: Visualizing pressure field around an airfoil

simpler. We only need to visualize the airfoil. One example is demonstrated in Figure 4a.

2.3 Design Choices

The performance of the model is highly sensitive to the choice of in-context examples, i.e., prompts (Zhang et al., 2023). It is of significance to experiment with different ways of prompting to generate desired outputs. Thus, before feeding the prompts into the inpainting model, we experiment with variations of visualization of airfoils and their surrounding pressure fields. For example, we can consider experimenting with different colors for both the airfoil and the pressure field, different sizes, and rotation.

2.3.1 Task Input Example: Only Airfoil

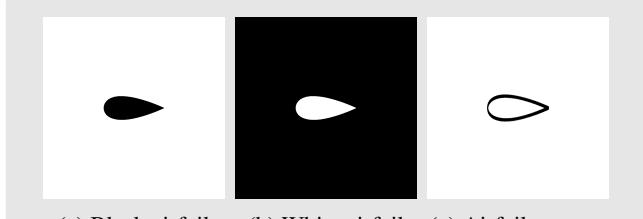
Intuitively, when plotting only the airfoil, our primary goal is to ensure that it is as distinct as possible from the rest of the image. To achieve this, we come up with three ideas only using black and white colors:

- Plotting the entire airfoil black against a white background (Figure 5a)
- Plotting the entire airfoil white against a black background (Figure 5b)
- Only plotting the contour of the airfoil black and leave the background white (Figure 5c)

After initial experiments, we conclude that the first approach is the best, i.e., black airfoil with white background, and we will stick with this one for the rest of this work.

2.3.2 Task Output Example: Airfoil and Its Surrounding Pressure Field

Color Should we opt for black-and-white images or colored images? Is it more effective to use diverging



(a) Black airfoil

(b) White airfoil

(c) Airfoil contour

Figure 5: Three ways of plotting only the airfoil

colors or different shades of a single color? To address these questions, we conduct extensive experiments with various color schemes. We defined a start color and an end color for each plot, assessing their impact on the clarity and effectiveness of our visual representations.

Some selected plots are shown in Figure 6. Figure 6a is the black-and-white version. The brighter regions indicate lower pressure coefficient values, while the darker ones correspond to higher values. Figures 6b, 6c, and 6e are similar to Figure 6a, only that they are colored versions. These figures have the same start color, i.e., white, and they display the pressure field using different shades of a single color.

It is also a possibility to set the start color black (Figure 6f). One may compare Figure 6b and Figure 6f. In this case, the darker regions indicate lower pressure coefficient values and brighter regions higher. We prefer setting the start color white. This approach makes the black airfoil more distinct against its background, thereby maintaining its shape more effectively when fed into the model. It should be noted that here we color all the airfoils black. This is also a design choice considering that we chose to plot the airfoil black against a white background for task input examples. This ensures consistency on airfoils between the task input and output examples. The black color also effectively distinguishes the airfoil from its surroundings.

Moreover, Figures 6d, 6g, and 6h are examples of using diverging colors for the pressure field. After initial experiments with these diverging colors, we conclude that using a single color yields better performance. Intuitively, using a single color allows the background's brightness to more effectively convey the degree of the pressure field, as compared to the use of two diverging colors.

Zoom-in Level The zoom-in level of the airfoil could also have a significant impact on the model results. We can investigate whether this is true by plotting the same airfoil with different zoom-in levels while keeping other variables constant. Figure 7 displays plots of the same

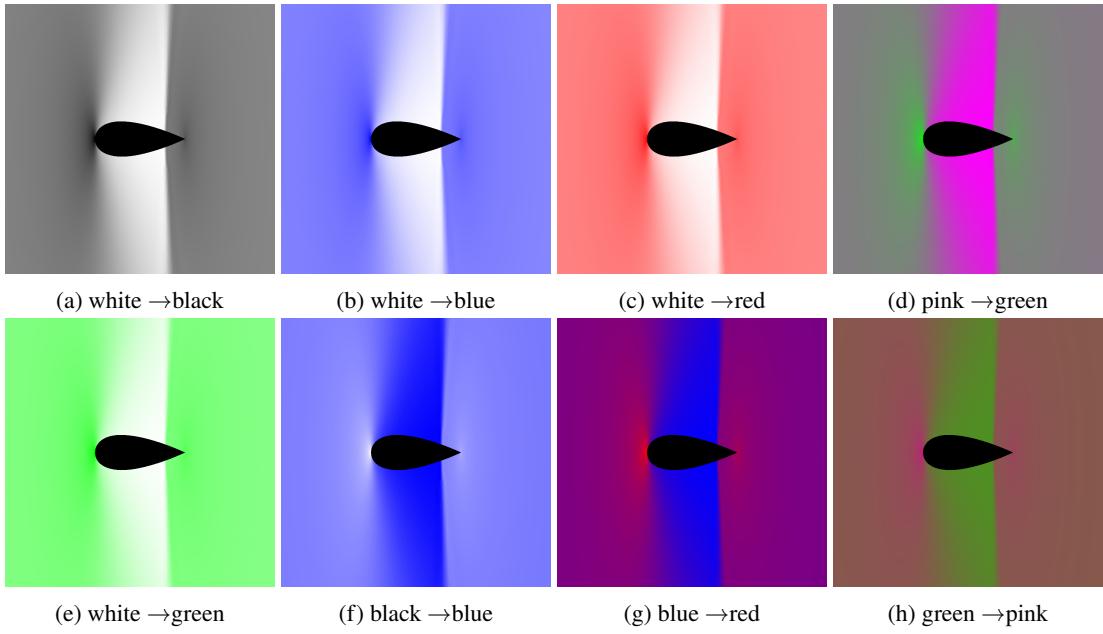


Figure 6: **Visualizations of an airfoil with surrounding pressure field of different colors.** For each image, the color stated before the arrow is the start color and the color after the arrow is the end color.

airfoil at three different zoom-in levels: small, medium, and large, while keeping other factors constant. We will conduct further research on the effect of airfoil zoom-in level later in this study.

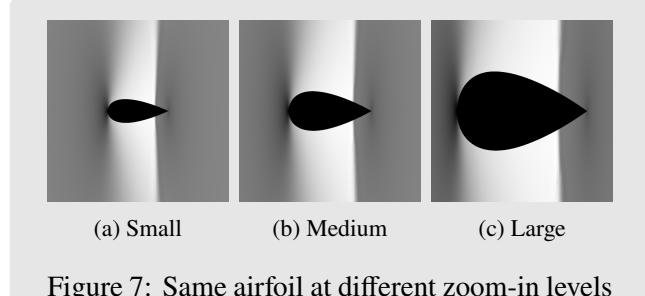


Figure 7: Same airfoil at different zoom-in levels

Rotation Would rotation help with the model results? Figures 8b and 8c are the rotated versions of Figure 8a. However, the model performs poorly with these rotated versions (see Figure 9). Thus, we will not consider rotation in our further analysis.

2.4 Measuring Metrics

How do we evaluate the prediction images generated by the model? With access to the pressure coefficient data for airfoil B, we can generate the actual image of airfoil B and its surrounding pressure field. We can compare this actual image with the model-generated image. The goal is to quantitatively measure:

- how well the airfoil shape is maintained

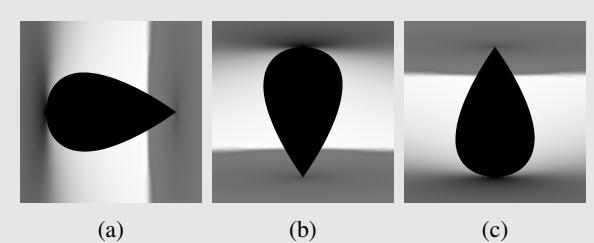


Figure 8: Same airfoil of different orientation

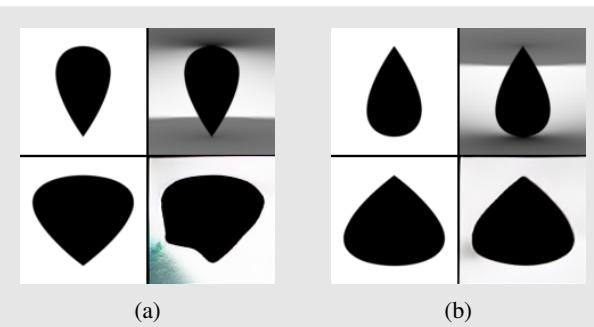


Figure 9: Model results for rotated prompts

- how accurately the pressure field is predicted.

In the following section, we introduce some quantitative metrics designed to evaluate the efficacy and precision of the model’s predictions:

Intersection over Union (IoU) The first metric we employ to assess the model’s accuracy is the *Intersection over Union (IoU)* of the airfoil. IoU is commonly used in image segmentation and object detection. It calculates the overlap between the predicted airfoil and the actual airfoil. The numerator is the intersection of the two airfoils and the denominator is the union of the two airfoils. This metric is effective in measuring how well the model preserves the geometric integrity of the airfoil’s shape.

Mean Squared Error (MSE) Another important metric is the *Mean Squared Error (MSE)*. This measure quantifies the average squared difference between the actual image and the predicted image, offering us a way to measure the difference between the actual pressure field around airfoil B and the predicted one. A low MSE indicates a close match between the actual image and the predicted one. This metric provides us with a clear and effective measure of the model’s performance in predicting the pressure field around an airfoil.

Structural Similarity Index (SSIM) Another metric is *Structural Similarity Index (SSIM)*, which is an image quality index. It is computed with respect to a reference image and is commonly used for measuring the similarity between two images. This measure takes into account three parameters: luminance, contrast, and structural information (Wang et al., 2004). It offers a more comprehensive and human-visual-system-aligned assessment. It provides another insightful perspective in measuring the model’s ability to predict the pressure field around an airfoil and preserve structural details.

SSIM ranges between -1 and 1. $SSIM = 1$ means perfect similarity and that two images are the same. $SSIM = 0$ means no correlation between the images. An SSIM value close to one would be desirable for our task.

Coefficient of Determination R^2 The R^2 value is a key metric in our evaluation. R^2 is the proportion of variation in the dependent variable that is predictable from the independent variables. We can also use it for image comparison. We use R^2 to determine how closely the model’s predictions of the pressure field around the airfoil align with the actual data.

The R^2 value ranges from 0 to 1. A R^2 value close to 1 indicates that the model provides a good fit to data; in our case, this means that the predicted image closely aligns with the actual image.

Overlay Difference To compare the predicted image and the actual image, we can attempt to visualize the difference between them using overlay difference. It provides a intuitive way to assess the accuracy of the

predictions. This visual comparison allows us to identify areas where the model performs well and areas where the discrepancies are most pronounced.

3 Experiments and Results - Prompt Engineering

In this section, we comprehensively and systematically explore various ways of prompting and investigate what kind of prompts achieve the best results. Through a series of experiments, we aim to establish a set of best practices in prompt engineering that can enhance the accuracy and efficiency of the model.

3.1 Airfoil Zoom-in Level

3.1.1 Initial Experiment

To examine the effect of airfoil zoom-in level on the model’s outcome, first, we conduct some preliminary experiments to obtain initial insights into how variations in airfoil zoom-in level influence the model’s performance.

We randomly select two airfoils, say A and B. We use airfoil A as the task example to predict the pressure field around airfoil B. We repeat the experiment three times, each time varying only the zoom-in level of the airfoils. The zoom-in levels tested are: small, medium, and large. Note that with each individual experiment, airfoils A and B are of the same size. The resulting grids are shown in Figure 10. The images on the bottom right of the grid are the predicted outcomes. For comparison, Figure 11 shows the actual airfoil B illustrated in different sizes with its surrounding pressure field.

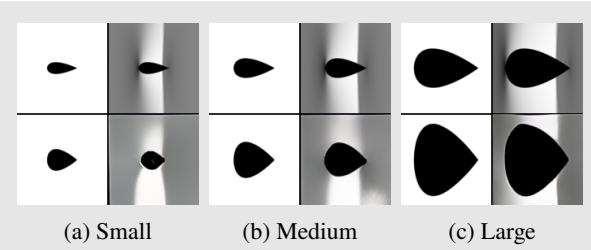


Figure 10: Model results for different airfoil zoom-in levels

We first inspect the difference between the predicted images and the actual images using visual overlay difference comparison (see Figure 12). Black areas indicate that there is little difference while brighter areas indicate larger discrepancies. As the airfoil zoom-in level increases, the shape of the airfoil is maintained better.

Next, we employ the quantitative metrics previously discussed. As depicted in Figure 13 and Table 1, the airfoil *IoU* increases as the airfoil zoom-in level increases,

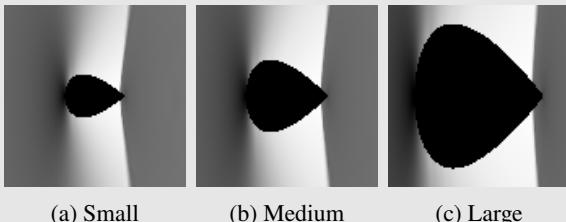


Figure 11: Actual airfoil B of different zoom-in levels

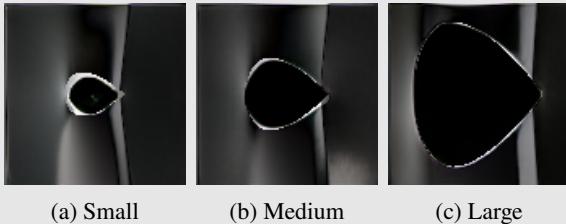


Figure 12: Overlay difference of the predicted and actual images

which is coherent with what we visually observed. The *MSE* decreases with increasing airfoil zoom-in level, indicating improving model performance. Both *SSIM* and R^2 values show an upward trend as the airfoil zoom-in level increases, further proving the positive impact of larger airfoil zoom-in level on model accuracy.

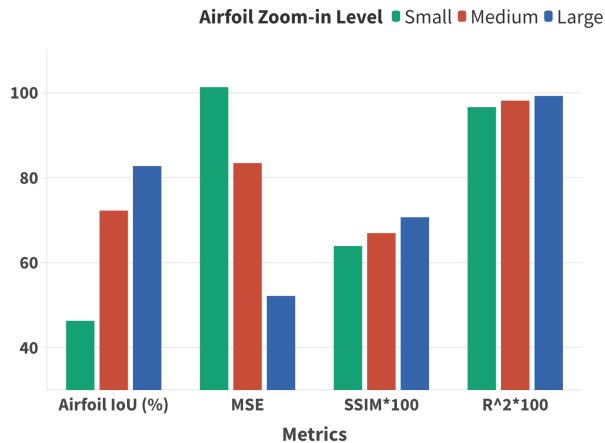


Figure 13: Quantitative metrics for different airfoil zoom-in levels

3.1.2 Repeated Experiments and Validation

To ensure the reliability and reproducibility of our finding, we replicate the initial experiment multiple times. This approach allows us to assess the consistency across

	Small	Medium	Large
Airfoil IoU	46.30%	72.25%	82.73%
MSE	101.31	83.44	52.16
SSIM	0.6390	0.6695	0.7068
R^2	0.9661	0.9814	0.9924

Table 1: Quantitative metrics for different airfoil zoom-in levels

multiple trials. It provides a more robust evaluation of the effect of airfoil zoom-in level on the model’s performance. We randomly select 20 airfoils to be airfoil A and another 20 airfoils to be airfoil B. Each replication follows the same procedure as the initial experiment.

Now, we can compare their metrics. To assess airfoil overlapping, we have three sets of *IoU* values corresponding to the small, medium, and large zoom-in levels of the airfoils. We call them *IoU_small*, *IoU_medium*, and *IoU_large*. Based on the initial experiment, we hypothesize that the *IoU* would be increasing as the airfoil zoom-in level increases. To verify this, we compare *IoU_small*, *IoU_medium*, and *IoU_large*. The proportion of increase from *IoU_small* to *IoU_medium* is 85%. This proportion is also 85% from *IoU_medium* to *IoU_large*. Furthermore, we conduct a t-test to investigate whether these sets of data are significantly different. The statistics are summarised in Table 2. For both transitions, from *IoU_small* to *IoU_medium* and from *IoU_medium* to *IoU_large*, the t-statistics are negative and the p-values are much smaller than 0.05. This indicates that the *IoU* values are increasing as airfoil zoom-in level increases and the differences are significant at 0.05 level.

For *MSE*, we also have three sets of *MSE* values corresponding to small, medium, and large sizes of the airfoils, namely *MSE_small*, *MSE_medium*, and *MSE_large*. Based on the initial experiment, we expect the *MSE* values to decrease as airfoil zoom-in level increases. The proportion of decrease from *MSE_small* to *MSE_medium* is only 50%, which is not particularly substantial. However, the proportion of decrease from *MSE_medium* to *MSE_large* is 90%, which is quite large. The *t-test* shows that the decrease from *MSE_medium* to *MSE_large* is statistically significant, whereas the decrease from *MSE_small* to *MSE_medium* is not (as detailed in Table 2).

For R^2 , the proportion of increase is 100% for both transitions: from small to medium and from medium to large. The t-test shows that the increases are highly significant (Table 2).

To conclude, larger airfoil zoom-in level does seem to

Metrics	Statistics	S to M	M to L
IoU	Proportion of increase	85%	85%
	t-stat	-4.4768	-3.4258
	p-value	0.0003	0.0028
MSE	Proportion of decrease	50%	90%
	t-stat	1.0572	6.9695
	p-value	0.3037	0.0000
R^2	Proportion of increase	100%	100%
	t-stat	-4.4655	-4.7435
	p-value	0.0003	0.0001

Table 2: Statistical tests for different airfoil zoom-in levels

enhance model accuracy.

3.2 Pressure Field Color

3.2.1 Initial Experiment

Same strategy as before, to investigate the impact of background color on model performance, we conduct some preliminary experiments to see whether we should use a colored background or just a black-and-white version.

We randomly select two airfoils A and B where A is the task example and B is the query. We repeat the prediction twice, once using a black-and-white background and once using a blue background. Figure 14 illustrates predicted images and the actual images for these two versions. We can see that in the lower right image of the black-and-white grid, there is an apparent discrepancy: a white area in the lower right corner where a darker tone is expected. In contrast, the blue version seems to be better. Figure 15 summarises the metrics comparing the black-and-white and blue background. R^2 is basically the same for both scenarios. The rest of the metrics demonstrate that the colored version is superior than the black-and-white version.

3.2.2 Repeated Experiments and Validation

Based on the initial experiments, we proceed to validate the findings through a series of repeated experiments. We attempt to ascertain the consistency and robustness of the metrics. We will also experiment with different colors and observe the results. We randomly choose 25 airfoils as airfoil A, and choose another 25 airfoils as airfoil B. We repeat the experiment 25 times, each time using one airfoil A and one airfoil B. This method provides a diverse and comprehensive examination of the model’s performance across a variety of airfoil combinations. For each experiment, we employ a range of colors for analysis, specifically black-and-white, blue, green, yellow, and red. In Figure 16, we display the model results for one experiment.

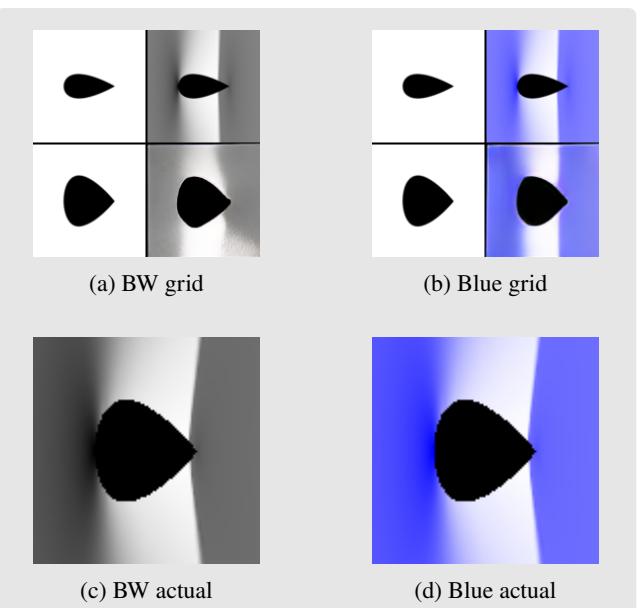


Figure 14: Black-and-white vs blue pressure field comparison

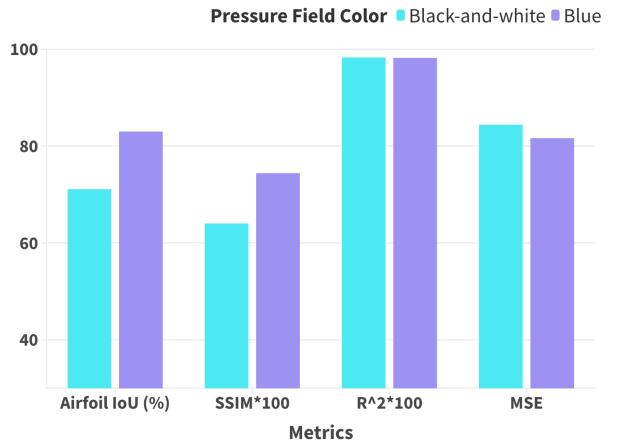


Figure 15: Quantitative metrics for BW and blue background

In Table 3, we summarise the statistical test results for experiments conducted using various background colors. First, we compare a few colored background with a standard black-and-white one. Then, we perform pair-by-pair comparisons among the colored backgrounds to see if there is any significant difference. As shown in the table, in most cases, the colored versions have better performance than the black-and-white version, except for a few exceptions. For the MSE of the transition from black-and-white to red, black-and-white background has better performance. For the R^2 of the transitions from black-and-white to blue and to red, the differences are not significant. Overall, considering the metric *IoU*, the

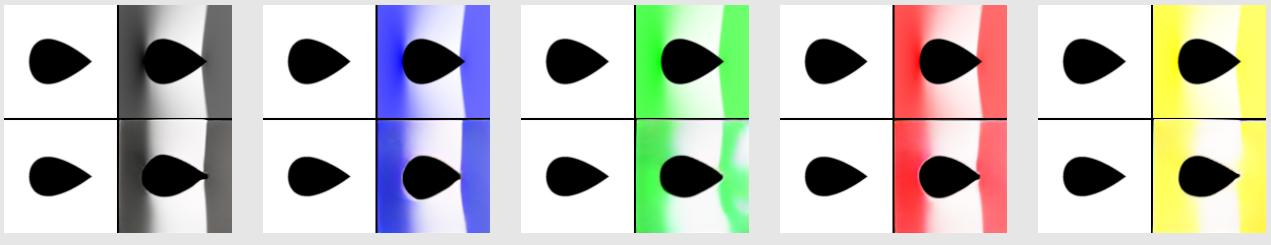


Figure 16: Model outcomes for various background colors

green background demonstrates the best performance, which means it is very good at maintaining the airfoil shape. For both metrics MSE and R^2 , the yellow color has the best performance, indicating superior model performance in predicting the pressure field when this color is used.

3.3 Aggregation

When using airfoil A to predict airfoil B’s pressure field, what if we repeat the same experiment multiple times and aggregate the outcome image? Would this approach yield a more accurate prediction? We choose airfoil NACA-0027 as the main subject, i.e., airfoil B. Next, we randomly select another 50 airfoils as airfoil A to predict airfoil NACA-0027. In other words, we replicate the experiment 50 times, with each replication involving a different airfoil A predicting the same airfoil B. To obtain the aggregation of these experiments, we compute the mean of the model result images (the bottom-right image of the grid). Then we can compare this aggregated image with the image from one single experiment. This comparison aims to assess whether the aggregation improves the accuracy of the prediction.

Figure 17a shows the overlay difference of a single result with the actual airfoil NACA-0027. Figure 17b is the overlay difference of the aggregated image with the actual airfoil NACA-0027. Comparing Figures 17a and 17b, visually, the aggregated version is better. The pressure field is darker, indicating more accurate prediction. Additionally, the outline of the airfoil is darker, meaning less difference with the actual airfoil. Moreover, Figure 18 shows the comparison of quantitative metrics for the aggregated result and the single result. Aggregated result is better for all metrics except for airfoil IoU .

3.4 Using Other Airfoils’ Pressure Coefficients

In our datasets, the number of points is the same for each airfoil. These points have their own point ID and each

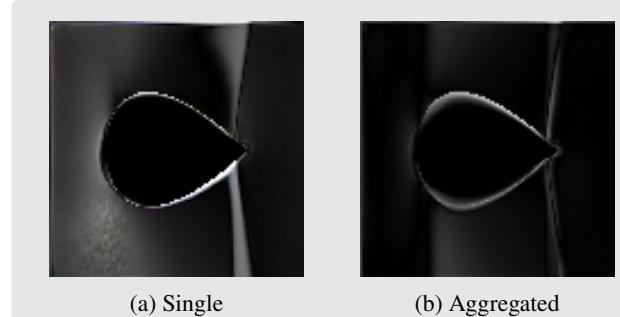


Figure 17: Overlay difference for aggregated result vs single result

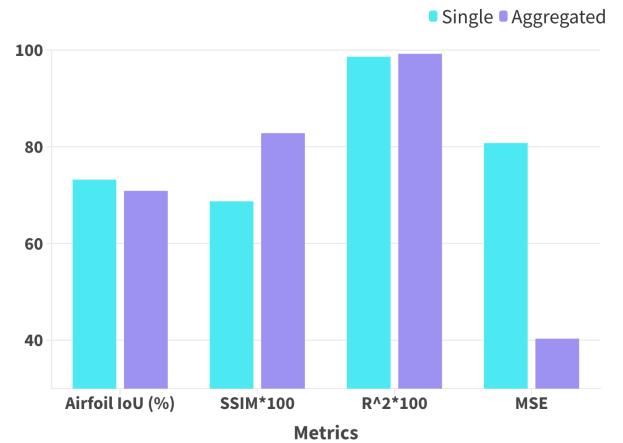


Figure 18: Quantitative metrics for aggregated result vs single result

point corresponds to a pressure coefficient value. Given a point ID, we know its coordinates and its pressure coefficient. Say that we do not know the pressure coefficients for airfoil B and know the pressure coefficients for airfoil C, to obtain the pressure field for airfoil B, can we use airfoil B’s coordinates and airfoil C’s pressure coefficient values? Would this result be better than using the inpainting model?

Metrics	Statistics	BW to Blue	BW to Red	BW to Yellow	BW to Green	Blue to Red	Red to Yellow	Green to Yellow
IoU	Proportion of increase	84.00%	96.00%	96.00%	100.00%	80.00%	76.00%	16.00%
	t-stat	-3.8249	-6.1465	-7.0184	-9.6985	-4.1433	-3.0957	3.7011
	p-value	0.0008	0.0000	0.0000	0.0000	0.0004	0.0049	0.0011
	Which is better?	Blue	Red	Yellow	Green	Red	Yellow	Green (best)
		BW to Blue	BW to Red	BW to Yellow	BW to Green	Red to Blue	Blue to Yellow	Green to Yellow
MSE	Proportion of decrease	88.00%	24.00%	100.00%	80.00%	96.00%	100.00%	96.00%
	t-stat	5.0574	-2.2113	12.1566	4.2700	7.0755	-12.5463	-9.8027
	p-value	0.0000	0.0368	0.0000	0.0003	0.0000	0.0000	0.0000
	Which is better?	Blue	BW	Yellow	Green	Blue	Yellow	Yellow (best)
R^2	Proportion of increase	28.00%	44.00%	100.00%	60.00%	28.00%	100.00%	100.00%
	t-stat	1.7332	1.5043	-19.7390	-3.2752	0.5066	-32.6433	-22.5735
	p-value	0.0959	0.1456	0.0000	0.0032	0.6171	0.0000	0.0000
	Which is better?	NA	NA	Yellow	Green	NA	Yellow	Yellow (best)

Table 3: Statistical tests for different background colors

Our main subject is still airfoil NACA-0027. To reach a more robust conclusion, we repeat the above method 50 times and aggregate the result. That is to say, in each experiment, we use one airfoil’s pressure coefficient and airfoil B’s coordinates to obtain airfoil B’s pressure field image. We use 50 different airfoils as airfoil C. The aim is to generate 50 separate predictions of airfoil B’s pressure field, which we then aggregate into a single composite result.

In [3.3], we have previously aggregated results from 50 repetitions of the experiment using the inpainting model. Now we can compare the aggregated results from these two methods. It should be noted that, for the method of using other airfoils’ pressure field, airfoil B’s shape would be perfectly maintained since we are using airfoil B’s coordinates in the first place. Thus, there is no point in comparing the two methods’ *IoU*.

In Figure [19], we show the aggregated results of the two methods we discussed above. Figure [19a] is the actual pressure field surrounding airfoil NACA-0027. Figures [19c] and [19d] are the aggregated results for using the inpainting model and using other airfoils’ pressure values respectively. Figures [19e] and [19f] are the overlay differences for using the two methods. The results indicate that using the pressure coefficients from other airfoils for our predictions yields a very favorable overlay difference.

Figure [20] provides a quantitative comparison between the two models. The method using other airfoils’ pressure coefficients shows superior performance. However, this could be due to the fact that we are using airfoil B’s coordinates in the first place. This inherent alignment of coordinates poses an intrinsic advantage to this method, potentially skewing the comparison in its favor.

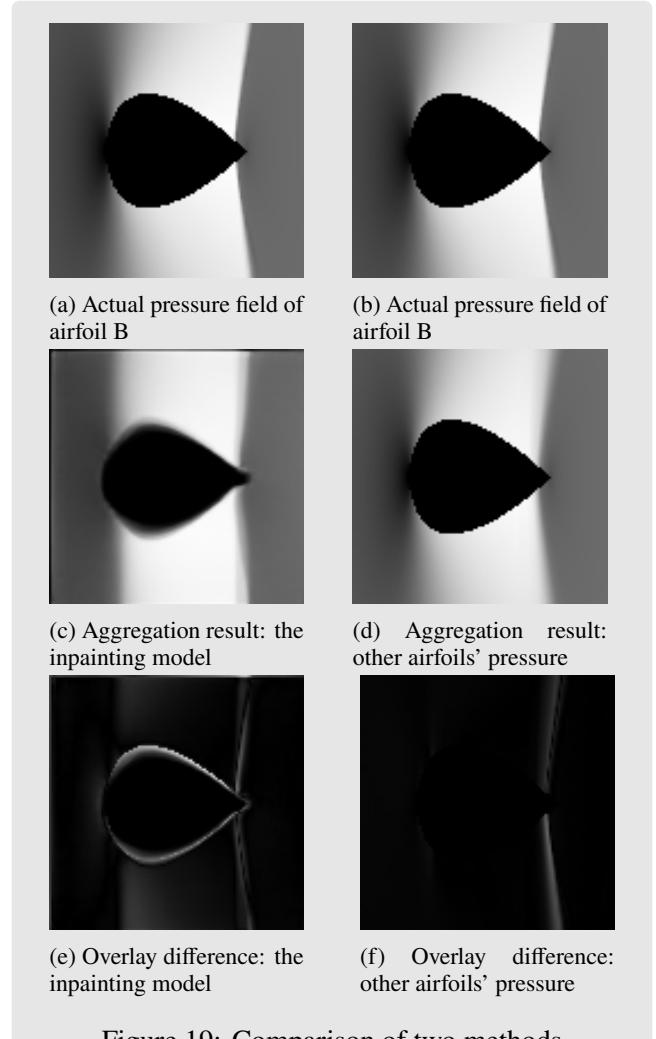


Figure 19: Comparison of two methods

3.5 Use Similar Airfoils or Dissimilar Airfoils as Prompts?

When we are using airfoil A to predict airfoil B’s pressure field, a natural question occurs: Is it more advantageous

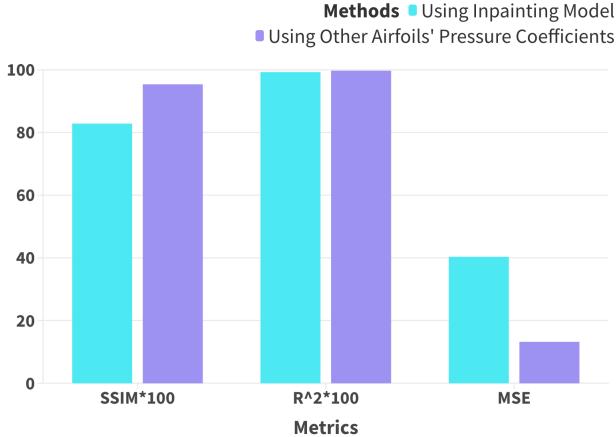


Figure 20: Quantitative metrics for comparing two methods

to select an airfoil A that is similar to airfoil B, or should dissimilarity be preferred? To investigate this, we still use airfoil NACA-0027 as our main subject, i.e., airfoil B. We select 10 airfoils that look very similar to airfoil NACA-0027 for prediction. We also select 10 airfoils that are very dissimilar structurally for prediction. Figure 21 demonstrates some of the airfoils that are very different from airfoil NACA-0027 structurally. Basically, we repeat the task of predicting the pressure field around airfoil B 20 times using similar and dissimilar airfoil As and compare their results.

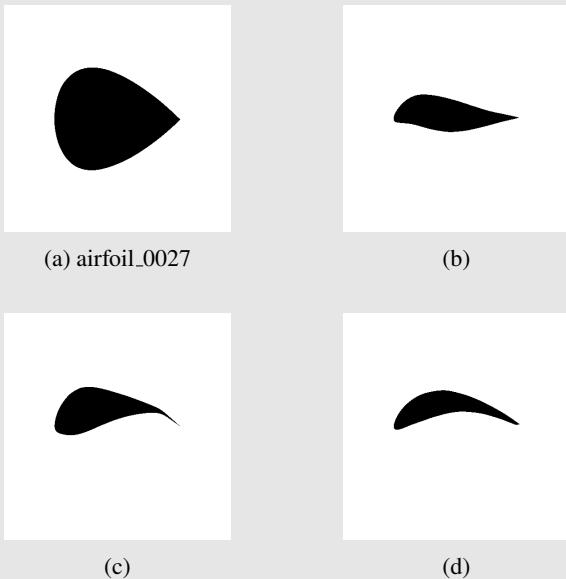


Figure 21: Main airfoil and several dissimilar airfoils

Figure 22 shows the model results for one example using a similar airfoil and one example using a dissimilar

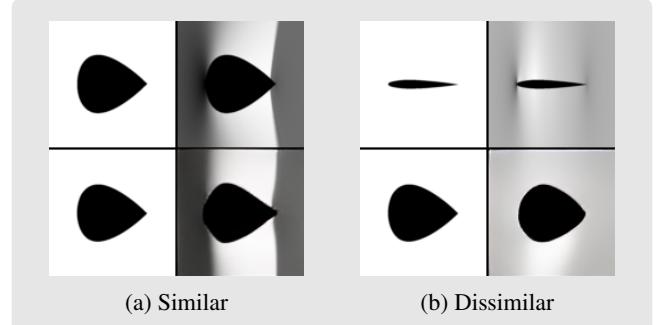


Figure 22: Model result comparison for using similar and dissimilar airfoils

Metrics	Statistics	Similar	Dissimilar
IoU	Mean	77.3%	70.4%
	t-stat	4.2343	
	p-value	0.0022	
MSE	Mean	35.901	72.044
	t-stat	-12.2865	
	p-value	0.0000	
R²	Mean	0.991	0.987
	t-stat	13.0401	
	p-value	0.0000	

Table 4: Quantitative metrics for using similar airfoils and dissimilar airfoils

airfoil. Figure 22a is for using similar airfoils and the prediction performance is adequate. Conversely, Figure 22b is for using dissimilar airfoils. Although airfoil B retains its shape well, the predicted pressure field is almost white, indicating very poor performance.

Next, we compare the quantitative metrics for repeated experiments for similar airfoils and dissimilar airfoils. For each of the experiment we conducted, we obtain a set of statistics. We take the average of these statistics for the similar set and the dissimilar set respectively. For similar airfoils, the average *IoU* value is 77.3%. For dissimilar airfoils, the average *IoU* value is 70.4%. The *t-test* shows that the difference between these two sets are significant, meaning that using similar airfoils is of advantage. Furthermore, the mean *MSE* for using similar airfoils is much smaller than that using dissimilar ones. The mean *R²* for using similar airfoils is higher than that using dissimilar airfoils, further substantiating the superior performance of using similar airfoils for prediction.

4 Discussion and Conclusion

Further to our previous discussion on choosing similar or dissimilar airfoils for prediction, the paper by Zhang

et al. (2023) provides a more comprehensive analysis. The paper investigates the impact of in-context examples on computer vision and concludes that the model performance is very dependent on the choice of in-context examples. It offers two methods for automating the selection of in-context examples: (1) an unsupervised prompt retrieval method and (2) a supervised prompt retrieval method that maximises in-context learning performance (Zhang et al., 2023). Results show that the supervised method is superior to the unsupervised method. The reason behind this is that the examples found by the supervised method are more similar to the queries. This could explain the superior performance when we use similar airfoils for prediction over dissimilar airfoils.

Prior to this, the field of natural language processing has already discovered that selecting in-context examples significantly influences performance, as highlighted by Agrawal et al.(2023) and Liu et al. (2022). Furthermore, the construction of these in-context examples (prompts), plays a crucial role in determining their effectiveness. Factors such as prompt length and the order of in-context examples have been identified as influential (Agrawal et al., 2023). These insights have motivated the community to explore optimal strategies for selecting in-context examples for large language models, which also inspires related research in computer vision.

In this work, we made an effort to predict the pressure field around airfoils using visual prompting. We experimented with various ways of prompting and compared their performance. However, this is only an initial attempt. There are still many challenges to be addressed. For instance, in the predicted image, the airfoil shape is not perfectly retained. We need more accurate results. Additionally, the predicted pressure field is not perfectly accurate. However, it could serve as a reference for further refinement.

References

- [Agrawal et al.2023] Sweta Agrawal, Chunting Zhou, Mike Lewis, Luke Zettlemoyer, and Marjan Ghazvininejad. 2023. In-context examples selection for machine translation. *Findings of the Association for Computational Linguistics: ACL 2023*.
- [Bar et al.2022] Amir Bar, Yossi Gandelsman, Trevor Darrell, Amir Globerson, and Alexei Efros. 2022. Visual prompting via image inpainting. *Advances in Neural Information Processing Systems*, 35:25005–25017.
- [Esser et al.2021] Patrick Esser, Robin Rombach, and Bjorn Ommer. 2021. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883.
- [Forrester et al.2006] Alexander IJ Forrester, Neil W Bressloff, and Andy J Keane. 2006. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071):2177–2204.
- [Giannakoglou2002] KC Giannakoglou. 2002. Design of optimal aerodynamic shapes using stochastic optimisation methods and computational intelligence. *Progress in Aerospace sciences*, 38(1):43–76.
- [He et al.2022] Kaiming He, Xinlei Chen, Saining Xie, Yang-hao Li, Piotr Dollár, and Ross Girshick. 2022. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009.
- [Liu et al.2022] Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2022. What makes good in-context examples for gpt-3? *Proceedings of Deep Learning Inside Out (DeeLIo 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*.
- [Radford et al.2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.
- [Selig1996] Michael S. Selig, 1996. *UIUC airfoil data site*. Department of Aeronautical and Astronautical Engineering, University of Illinois at Urbana-Champaign.
- [Wang et al.2004] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- [Zhang et al.2023] Yuanhan Zhang, Kaiyang Zhou, and Ziwei Liu. 2023. What makes good examples for visual in-context learning? *arXiv preprint arXiv:2301.13670*.