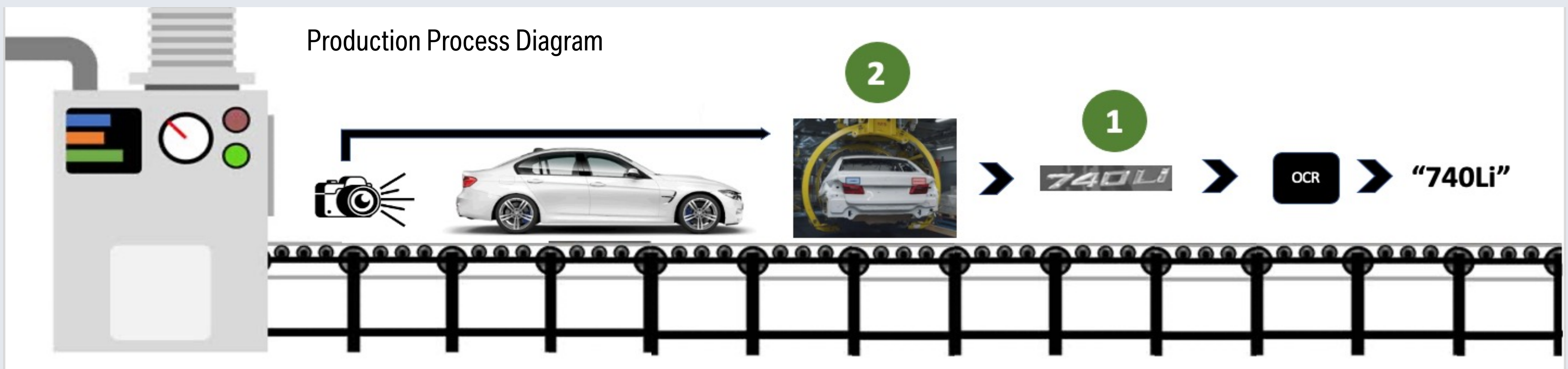
DEEP LEARNING FOR COMPUTER VISION: AUTOMATING QUALITY CONTROL IN CAR PRODUCTION

ANNITA VAPSI AND NONIE THOMAS
DR MAKA KARALASHVILI AND VASILEIOS MAGIOGLOU
PROF BART PAUL GERARD VAN PARYS



Acronym Key:
- OCR: Optical Character Recognition
- GAN: Generative Adversarial Network

References
[1] Sharon Fogel, Hadar Averbuch-Elor, Sarel Cohen, Shai Mazor, & Roee Litman. (2020). ScrabbleGAN: Semi-Supervised Varying Length Handwritten Text Generation.

[2] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaole Huang, & Xiaodong He. (2017). AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks.

INTRODUCTION

PROJECT 1: SYNTHETIC DATA GENERATION

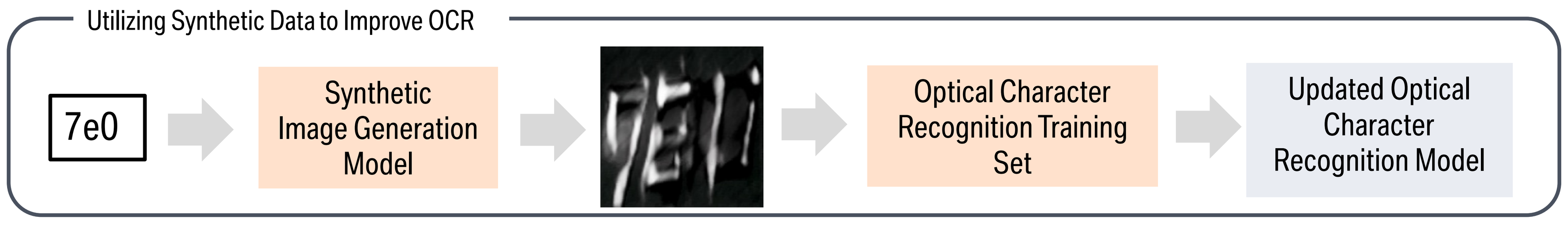
The model inscription images (label 1 in the production process diagram) are passed through an Optical Character Recognition (OCR) model, which recognizes the text in the image to ensure that the car body received the correct inscription. This model is very accurate (>99% accuracy), however it fails for images of model inscriptions which are new to the BMW fleet and were not included in the OCR training set. Our goal is to generate synthetic images of new model inscriptions to augment the training dataset and allow pre-scheduled retraining of the OCR model when new inscriptions are set to be introduced to the production line.

PROJECT 2: DATA DRIFT DETECTION

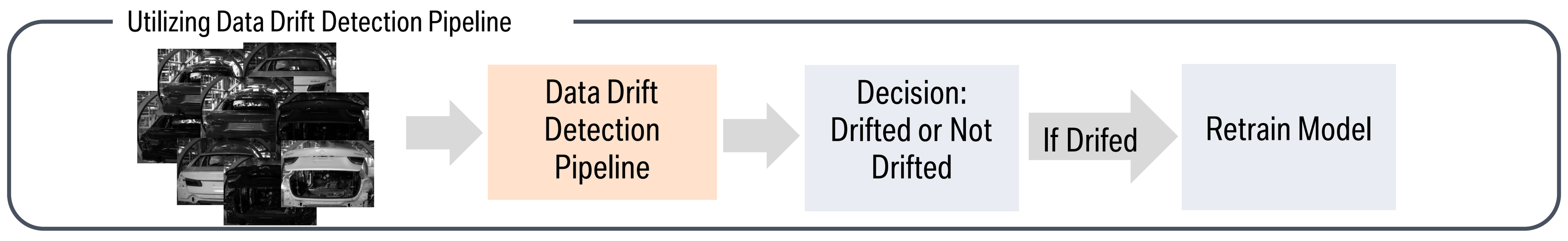
Over time, machine learning models can deteriorate in predictive power, a phenomenon called model drift. We focus on a subset of model drift called data drift, which occurs when there is a shift in the input data to a model. The aim of this project is to provide BMW Group with a methodology to assess when a model needs to be retrained due to data drift. The data set includes full rear images of cars (label 2 in the production process diagram above) including both drifted and undrifted images.

BUSINESS IMPACT

When new model inscriptions are introduced, BMW has to wait 2-4 weeks before enough model inscription image data can be collected to train the OCR model to a high enough accuracy. In the meantime, automated checks on model inscriptions are replaced by purely manual intervention. With our synthetic data, the OCR model could be trained on input data containing both the old and new inscriptions and utilized immediately, bypassing the need for manual checks which are time consuming and costly.

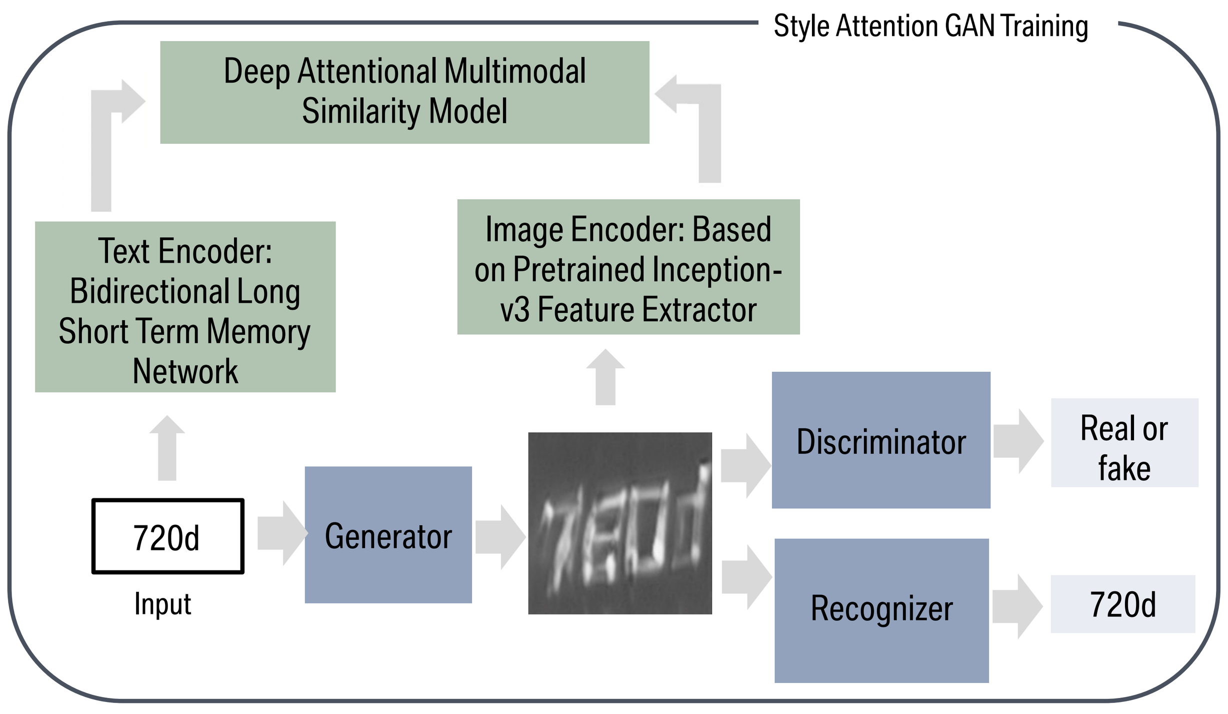


BMW leverages automation, including machine learning models, in several stages of the production and quality control of its car manufacturing process. Detecting this drift could prevent model accuracy deterioration, limit the need for manual quality control checks, and streamline model performance assessment.

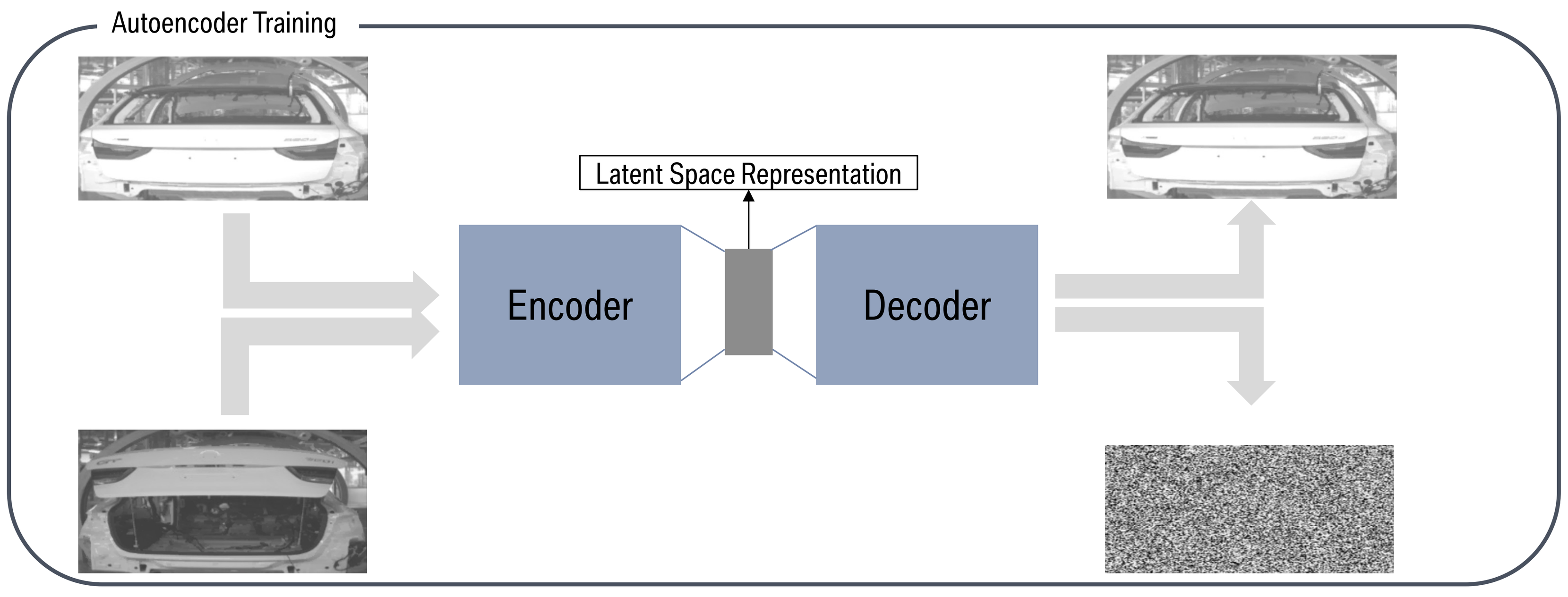


METHODS

We use a deep learning network, called a Generative Adversarial Network (GAN) to transform the textual model inscription (input), to a synthetic but realistic image. A GAN is composed of two competing agents: the generator and the discriminator. The generator aims to produce fake but realistic images while the discriminator aims to distinguish between real and fake images.



We used a Style Attention GAN [2] which leverages a pretrained attentional network, the Deep Attentional Multimodal Similarity Model to relate the textual input condition to the generated image using a text encoder and an image encoder. The image encoder is based on an Inception-v3 model which we pretrained. We included a component called the recognizer, which is a pretrained OCR model that penalizes the generator if the synthetic image is not readable or if the decoded inscription does not match the input condition [1].

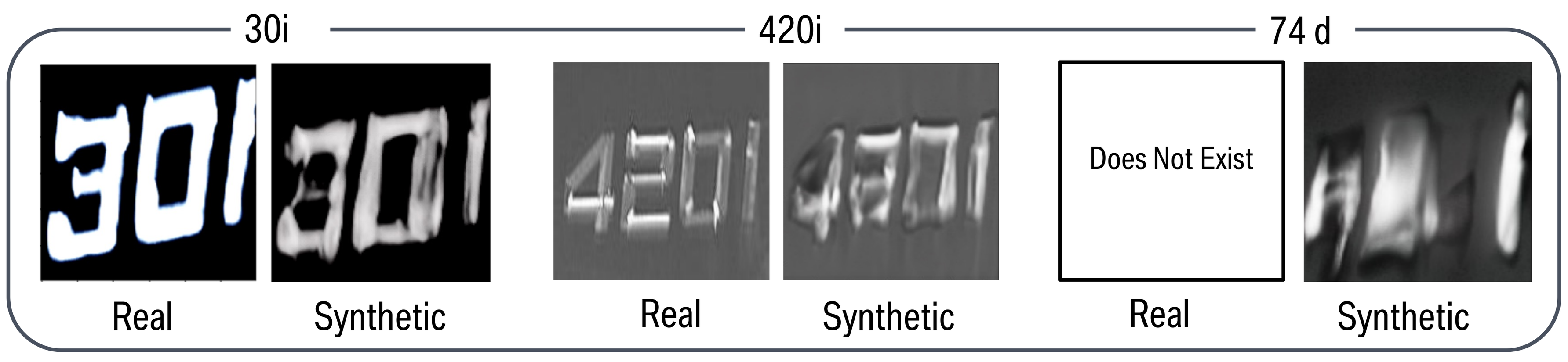


We use a neural network called an Autoencoder which aims to reproduce an input image exactly using encoder and decoder networks. Training the model on undrifted data, we leverage the model's inability to reconstruct data that has drifted. When the input data changes, the autoencoder is unable to reconstruct the image with the same precision.

We feed images from two datasets (one undrifted and one potentially drifted) through the trained Autoencoder and measure the reconstruction losses. We use the nonparametric Mann Whitney test to test whether the two sets of reconstruction losses are drawn from the same distribution. If they are not, the data has likely drifted.

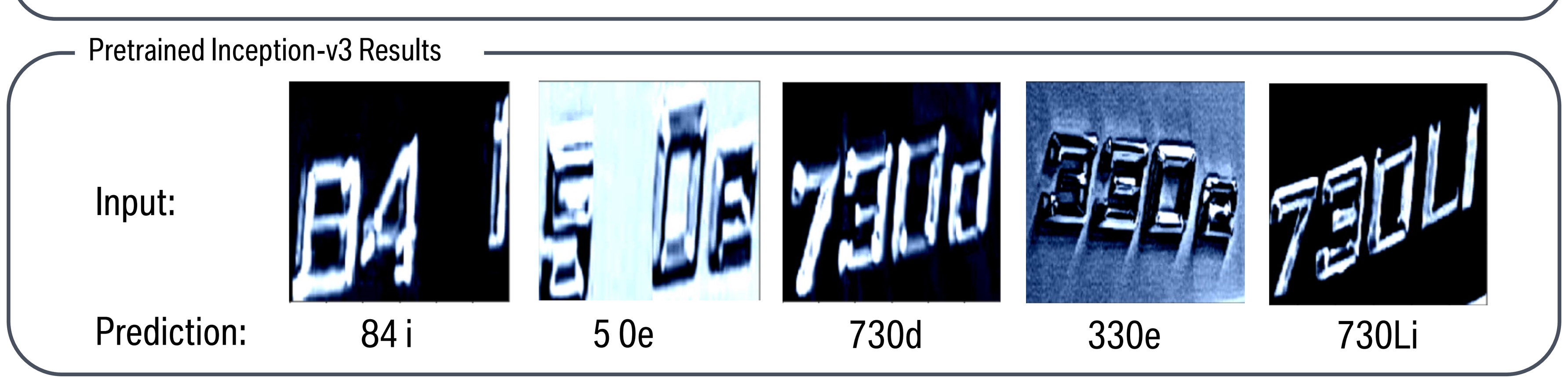
RESULTS

We visually inspect the ability of the model to generate images given an input model inscription as text. We test both inscriptions that exist in the dataset already (such as 30i and 420i) and those which are new and unseen by the model (such as 74 d).



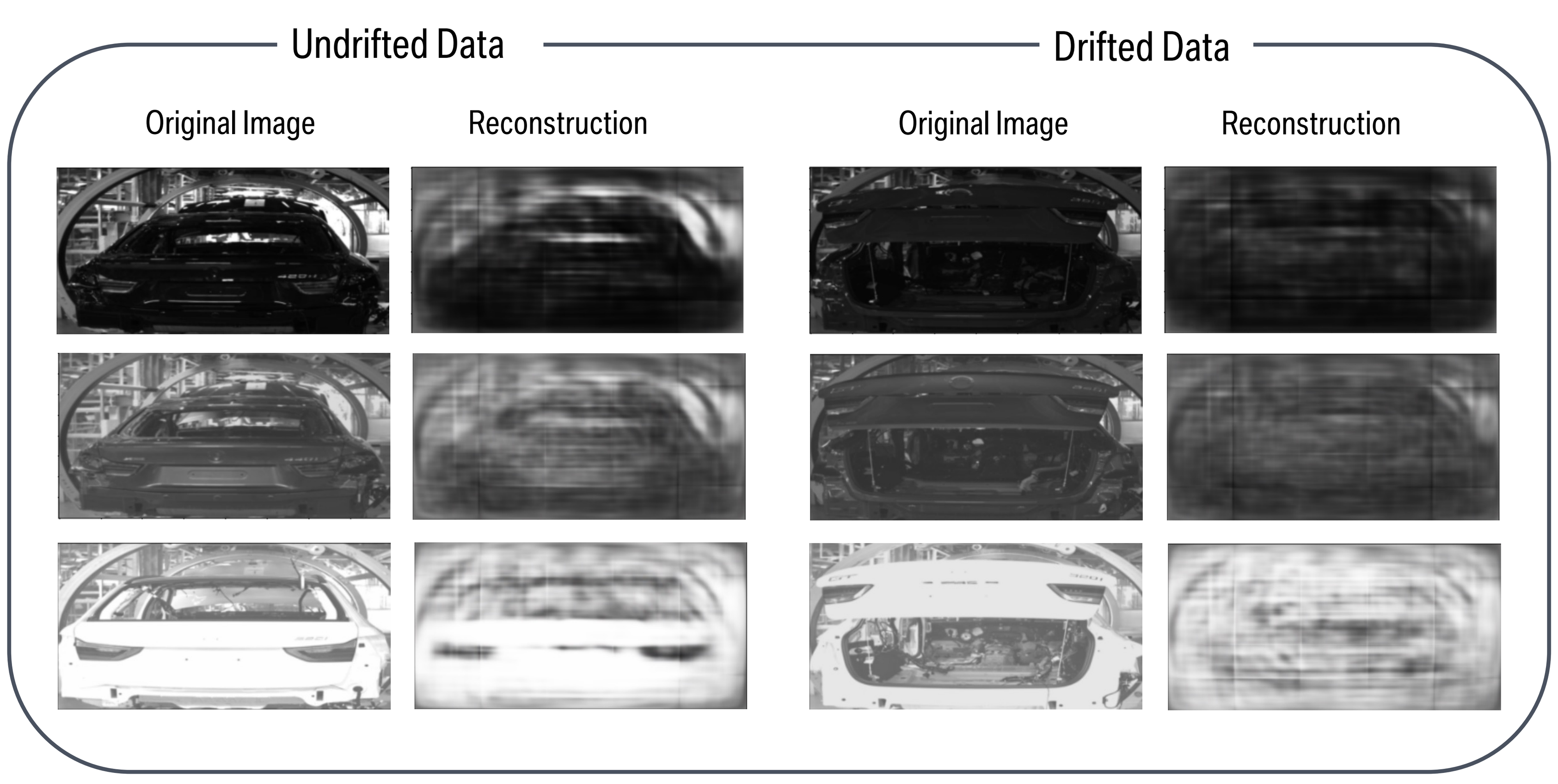
We evaluate the added benefit of the recognizer by comparing two datasets of synthetic images generated from inscriptions in the current BMW fleet using the Style Attention GAN, one including the recognizer component and the other excluding it.

- The Frechet Inception Distance compares the similarity of between real and generated datasets and aims to mimic human perception of similarity. It does not evaluate caption matching and a lower score is better. The addition of the recognizer improves this metric by a factor of 0.4.
- We obtain the accuracy of the OCR prediction on synthetic images. The addition of the recognizer improves this value by a multiple of 38.

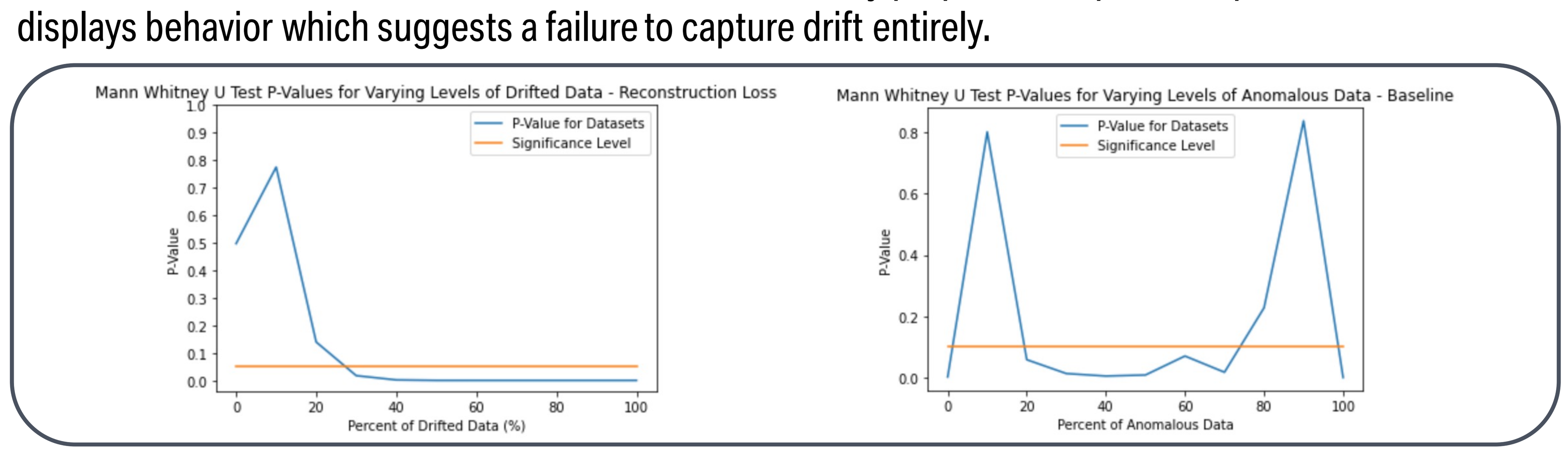


Key Takeaway: Our model with the recognizer is able to produce more visually realistic and readable images which better match their input condition than the baseline model without the recognizer. Additionally, it is somewhat able to capture new character orderings in unseen model inscriptions.

To visually assess whether there is a difference in performance when the autoencoder attempts to reconstruct images of undrifted data, we push through the model drifted and undrifted images independently.



We test the ability of our method, alongside the baseline method, to detect drift in datasets which are increasingly polluted by drifted data. The figure indicates that our method successfully detects data drift and is able to detect drift with a confidence of 95% for any proportion of pollution past 30%. The baseline displays behavior which suggests a failure to capture drift entirely.



ACKNOWLEDGMENTS

We would like to thank BMW Group for sponsoring this project and for the exciting collaboration as well as for hosting us in Munich this summer. We are grateful for the support of our BMW Group advisors, Dr. Maka Karalashvili and Vasileios Magioglou, and for the guidance of our faculty advisor Professor Van Parys.