

PRIMERJAVA DALL-E IN STABLE DIFFUSION ZA SINTETIČNO GENERIRANJE UČNIH PODATKOV ZA KLASIFIKACIJO SLIK Z MODELOM MOBILENETV3 IN NJEGOVA INTEGRACIJA V FLUTTER APLIKACIJO

Aleksandr Shishkov^{1,3}, Stevanče Nikoloski^{2,3}

¹ Univerza v Mariboru Fakulteta za elektrotehniko, računalništvo in informatiko, Koroška cesta 46, 2000 Maribor, Slovenija

² Univerza v Novem mestu Fakulteta za ekonomijo in informatiko, Na Loki 2, 8000 Novo mesto, Slovenija

³ Result d.o.o., Celovška cesta 182, 1000 Ljubljana, Slovenija, www.result.si

aleksandr.shishkov@student.um.si, stevance.nikoloski@result.si

Naša raziskava se osredotoča na izboljšanje klasifikacijskega modela za prepoznavanje slik z integracijo sintetično generiranih slik, ustvarjenih z uporabo dveh naprednih modelov text-to-image: Dall-E in Stable Diffusion. Naš cilj je ne le primerjati tehnološko napredne modele za sintetično generiranje slik, temveč tudi preučiti, kako lahko te slike obogatijo učno množico za model klasifikacije, zasnovan na MobileNetV3 arhitekturi, zlasti pri izboljšanju prepoznavanja robnih primerov. Ti robni primeri vključujejo različne svetlobne pogoje, kot so slike istega avtomobila ponoči ali čez dan, v dežju, snegu, megli in podobno, kar predstavlja dodaten izziv pri natančni klasifikaciji.

Ta pristop je še posebej relevanten za izboljšanje prepoznavanja specifičnih scenarijev, kot so nadzorne plošče avtomobilov, slike sprednjih in zadnjih delov avtomobilov ter dokumenti, kot so zavarovalne police in prometna dovoljenja. Sintetično generirane slike s pomočjo modelov Dall-E in Stable Diffusion namreč nadopolnjujejo in razširjajo obstoječi nabor podatkov, s čimer izboljšamo zmogljivost modela MobileNetV3 za natančno klasifikacijo slik v raznolikih pogojih.

Naša primerjalna analiza temelji na preizkušanju učinkovitosti modelov, ki so bili naučeni z dodatkom sintetično generiranih slik, v primerjavi z originalnim klasifikacijskim modelom, ki je dosegel visoko stopnjo natančnosti. Raziskujemo, kako izbira ustreznega prompta za generiranje slik vpliva na kvaliteto nabora podatkov in posledično na točnost klasifikacijskega modela. Prav tako preučujemo zmogljivosti modelov med treningom, validacijo in ročnim testiranjem, da ocenimo praktično uporabnost tako generiranih naborov podatkov.

Pomemben del naše raziskave predstavlja tudi integracija izboljšane modela MobileNetV3 v mobilno aplikacijo Flutter, kar uporabnikom omogoča neposredno uporabo naprednih klasifikacijskih modelov na mobilnih napravah. Ta integracija poudarja praktično vrednost našega dela, saj izboljšuje učinkovitost in prilagodljivost sistemov za slikovno analizo v realnem času, kar je ključnega pomena za aplikacije, ki zahtevajo hitro in zanesljivo prepoznavanje slikovnih podatkov v širokem spektru okoljskih pogojev.

Ključne besede: Dall-E, Stable Diffusion, text-to-image, MobileNetV3, klasifikacija slik, Flutter

COMPARISON OF DALL-E AND STABLE DIFFUSION FOR SYNTHETIC DATA GENERATION FOR IMAGE CLASSIFICATION WITH MOBILENETV3 MODEL AND INTEGRATION INTO FLUTTER APPLICATION

This article explores enhancing image classification models by integrating synthetically generated images using advanced text-to-image models: Dall-E and Stable Diffusion. Aimed at improving the accuracy of edge case recognition—such as different lighting conditions, images of the same car by day or night, under rain, snow, or fog—this approach significantly enriches the training dataset for a MobileNetV3-based classification model. By augmenting the original dataset with synthetic images, we aim to boost the model's ability to accurately classify specific scenarios, including car dashboards, front and rear car views, and documents like insurance policies and vehicle permits, thereby preventing the upload of irrelevant images into a Flutter-developed mobile application. Our comparative analysis evaluates the effectiveness of models trained with these enriched datasets against a baseline classifier, highlighting the practical utility of generated datasets and integrating an enhanced MobileNetV3 model into real-time image analysis systems within mobile environments.

Keywords: Dall-E, Stable Diffusion, text-to-image, MobileNetV3, image classification, Flutter.