



Neironu tīklu arhitektūras attēlu sintēzei no teksta

Attēlu sintēze no teksta

Attēlu sintēze no teksta ir samērā jauns uzdevums, kura mērķis ir no īsa tekstuāla apraksta ģenerēt tam atbilstošu un ticamu attēlu. Pēdējos gados šī uzdevuma risināšanā bijuši daudz mēģinājumu, galvenokārt pielietojot ģenerējošos tīklus-pretiniekus (*Generative Adversarial Networks* jeb GAN), taču nesen arī *transformer* tipa neironu tīklus. Uzdevums netiek atrisināts perfekti, taču rezultāti ir iespaidīgi.

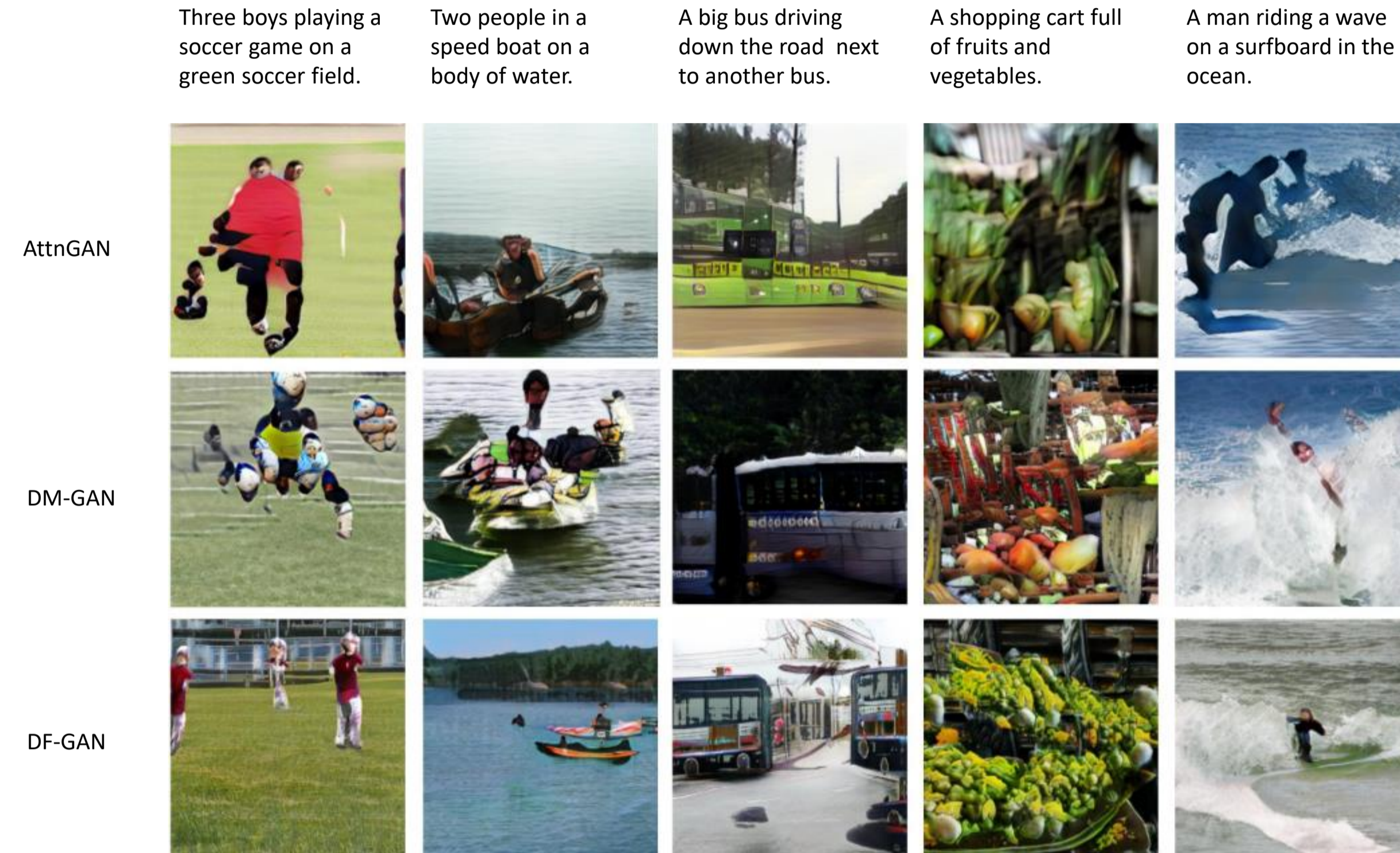
Darba mērķis

Noskaidrot, cik tālu attēlu sintēze no teksta ir attīstījusies, kādi ir galvenie virzieni šī uzdevuma risināšanā, kā arī izpētīt darbības principus labāk strādājošajiem risinājumiem.

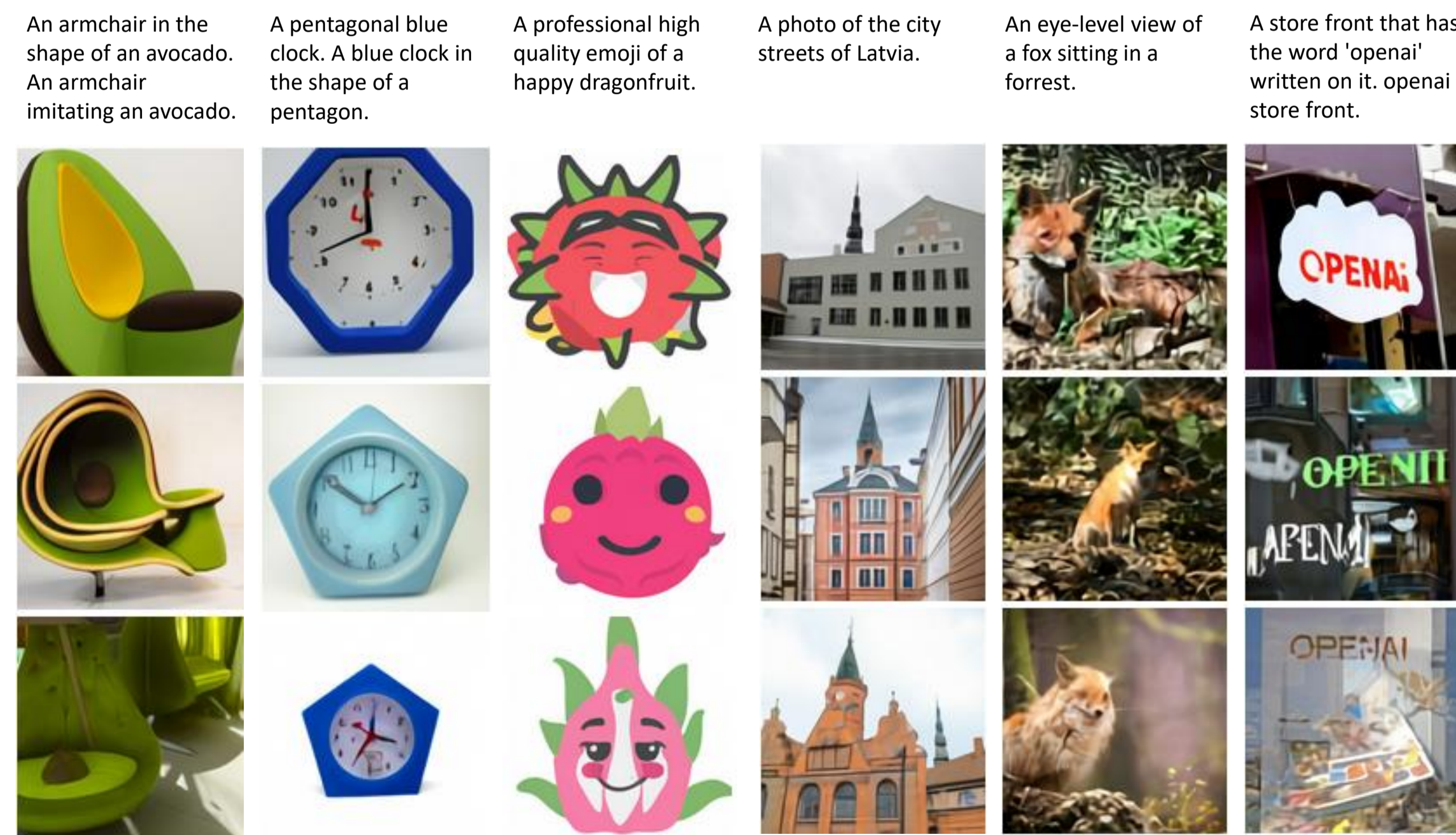
Rezultāti

- Apskatītas 3 GAN pamat-arhitektūras.
- 35 GAN arhitektūrām attēlu sintēzei no teksta noskaidrotas to izmantotās datu kopas, ģenerēto attēlu kvalitātes metrikas un rezultāti, kā arī koda pieejamība.
- 5 no 35 kvantitatīvi apskatītajām arhitektūrām pētīti to darbības principi.
- Apskatīts nesen iznākušais DALL-E risinājums un tā pamatā esošā CLIP modeļa darbības principi.
- Veikts apskats par 3 datu kopām, kuras literatūrā visbiežāk pielietotas apmācībai.
- 5 literatūrā populārākajām ģenerētu attēlu kvalitātes metrikām aprakstīti to darbības principi.

Dažādu arhitektūru GAN ģenerēti attēli



DALL-E ģenerēti attēli



Secinājumi

Izskatot pieejamo literatūru, darbā secināts, ka:

- gandrīz visiem risinājumiem pamatā ir GAN, taču nosacīti iespaidīgākajam un nesenākajam - *transformer* tipa neironu tīkli;
- uz šaurām datu kopām GAN spēj ģenerēt ļoti kvalitatīvus attēlus, taču uz reālistiski daudzveidīgām - līdz ideāliem rezultātiem vēl tālu;
- modeļu apmācība attēlu sintēzei no teksta ir ļoti resursietilpīga;
- ģenerēto attēlu kvalitātes metriku ir daudz, no populārākajām visām ir trūkumi, turklāt ar vairākām metrikām iespējams manipulēt, lai iegūtu labākus rezultātus. Bez tam, katrs risinājums tiek testēts ar atšķirīgu metriku kopu un pat vienas metrikas ietvaros, ģenerēto attēlu kvalitātes testēšanas metodoloģija var būtiski atšķirties starp publikācijām.

Turpmākie soļi

- Izskatīt iespējas attēlu sintēzei no teksta izmantot *transformer* tipa neironu tīklus, tai skaitā dziļāk izpētīt DALL-E, kad tā autori izdos solīto zinātnisko rakstu.
- Izpētīt iespēju CLIP modeli izmantot attēlu ģenerēšanai (nevis tikai rezultātu sarindošanai pēc atbilstības tekstam).
- Izveidot jaunu neironu tīklu arhitektūru attēlu sintēzei no teksta, kuru iespējams apmācīt saprātīgi īsā laika periodā ar vienu grafisko karti.