

Posudek vedoucího diplomové práce

Autorka práce: Bc. David Kunz

Název práce: *Text-driven Real-time Video Stylization Using Diffusion Models*

Vedoucí práce: prof. Ing. Daniel Sýkora, Ph.D., ČVUT FEL (K13139)

Úkolem diplomové práce David Kunze bylo seznámit se s existujícími metodami pro přenosu výtvarného stylu do videa, které byly původně navrženy pro práci s ručně kreslenými předlohami [3, 4, 2] a zaměřit se na možnost tvorby stylizovaných snímků s využitím textového vstupu pomocí difuzního modelu [1]. Cílem práce bylo navíc dosáhnout interaktivní odezvy a umožnit tak uživateli provádět stylizaci živého videa v reálném čase např. pro účely konferenčního videohovoru. Přestože se podobná řešení založená na využití difúzních modelů v současné době hojně objevují, žádné z nich nedokáže zachovat konzistenci výsledné stylizace v reálném čase a právě v tomto aspektu měla práce inovativní charakter.

Student se svého úkolu zhostil s velkou pílí a nadšením. Pravidelně konzultoval svůj postup s vedoucím práce a samostatně navrhoval další nová vylepšení, jež byla nad rámec původního zadání. Postupně se detailně seznámil s metodou umožňující stylizaci videa v reálném čase [4] a inovativním způsobem ji zkombinoval s přístupem založeným na použití difuzního modelu [1]. Ve výsledném řetězci následně implementoval řadu dalších zlepšení, která umožnila zvýšit celkovou konzistenci stylizace při zachování interaktivní odezvy. Vznikl tak interaktivní nástroj, který svými kvalitami předčí současné nejlepší pokusy o řešení podobného problému. Velmi oceňuji také studentovu snahu provést důkladné testování nově navrženého přístupu. Dokladem toho byl zejména interaktivní workshop organizovaný ve spolupráci se skupinou Uroboros, kde si účastníci mohli systém vyzkoušet osobně v praxi a poskytnout autorovi cenné připomínku a návrhy na další zlepšení. Díky těmto aspektům získala práce zjevný publikační potenciál, který se, jak doufám, podaří v brzké době realizovat.

Samotný text práce považují taktéž za zdařilý. Oceňuji fakt, že byl napsán v angličtině na slušné jazykové úrovni a může tak posloužit jako dobrý základ pro budoucí publikaci. Velmi cenný je v něm též názorný přehled současného stavu poznání v oblasti stylizace videosekvencí, který by mohl být sám o sobě dobrým základem pro případný přehledový článek. Další významné pozitivum vidím v rozsáhlé kapitole věnované prezentaci výsledků testování. Názorně se zde ukazuje praktický potenciál nově navrženého přístupu a jeho možné limitace.

S diplomovou prací Davida Kunze jsem velmi spokojen, doporučuji ji k obhajobě, hodnotím stupněm **A-výborně** a navrhuji komisi zvážit její nominaci na **cenu děkana**.

V Praze, 24. ledna 2024

prof. Ing. Daniel Sýkora, Ph.D.

Literatura

- [1] Brooks, T.; Holynski, A.; Efros, A. A.: InstructPix2Pix: Learning to Follow Image Editing Instructions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2023, s. 18392–18402.
- [2] Futschik, D.; Kučera, M.; Lukáč, M.; aj.: STALP: Style Transfer with Auxiliary Limited Pairing. *Computer Graphics Forum*, ročník 40, č. 2, 2021: s. 563–573.
- [3] Jamriška, O.; Šárka Sochorová; Texler, O.; aj.: Stylizing Video by Example. *ACM Transactions on Graphics*, ročník 38, č. 4, 2019: str. 107.
- [4] Texler, O.; Futschik, D.; Kučera, M.; aj.: Interactive Video Stylization Using Few-Shot Patch-Based Training. *ACM Transactions on Graphics*, ročník 39, č. 4, 2020: str. 73.