

Deep-Learning-basierte Analyse und Synthese von Gesichtern in verschiedenen Kunststilen

Anna Taphorn



Inhalt

1. Ziel der Arbeit
2. Motivation
3. Methodik - Analyse
 - i. Datensatz
 - ii. Semantische Labels
 - iii. Statistische Auswertung
 - iv. Geometrie der Gesichter
 - v. Farben
 - vi. Klassifizierung
 - vii. t-SNE
 - viii. Visualisierung
4. Methodik – Generierung
 1. Fragestellungen
 2. Regression
5. Literatur



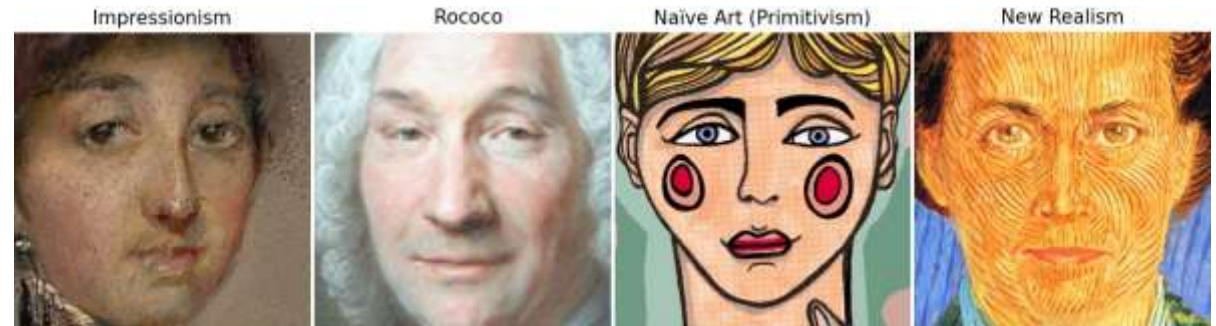
Ziele und Motivation



Ziel der Arbeit

Ziele:

1. Analyse von Gesichtern verschiedener Kunststile mithilfe von Deep-Learning-Algorithmen
2. Synthese von künstlichen Gesichtsbildern, im Stil spezifischer Kunstklassen



Forschungsfragen:

- Gibt es Unterschiede, Gemeinsamkeiten und Zusammenhänge zwischen Gesichtsdarstellungen in verschiedenen Kunststilen, die durch Deep-Learning-Algorithmen aufgedeckt werden können?
- Können die gewonnenen Erkenntnisse auf künstlich erzeugte Gesichter übertragen werden?

Motivation

- Quantifizierbare Analyse von Kunst
→ Mögliche kunstwissenschaftliche Erkenntnisse
- Grundlagenforschung zu generativen stateof-the-art Modellen
→ tieferes Verständnis für die Möglichkeiten und Grenzen von GANs

Warum Gesichter?

- Geben Aufschluss über Gesellschaft und Menschenbild
- Hohe Vergleichbarkeit
- In fast jedem Kunststil vorhanden
- Bisher wenig Forschung

Methodik - Analyse

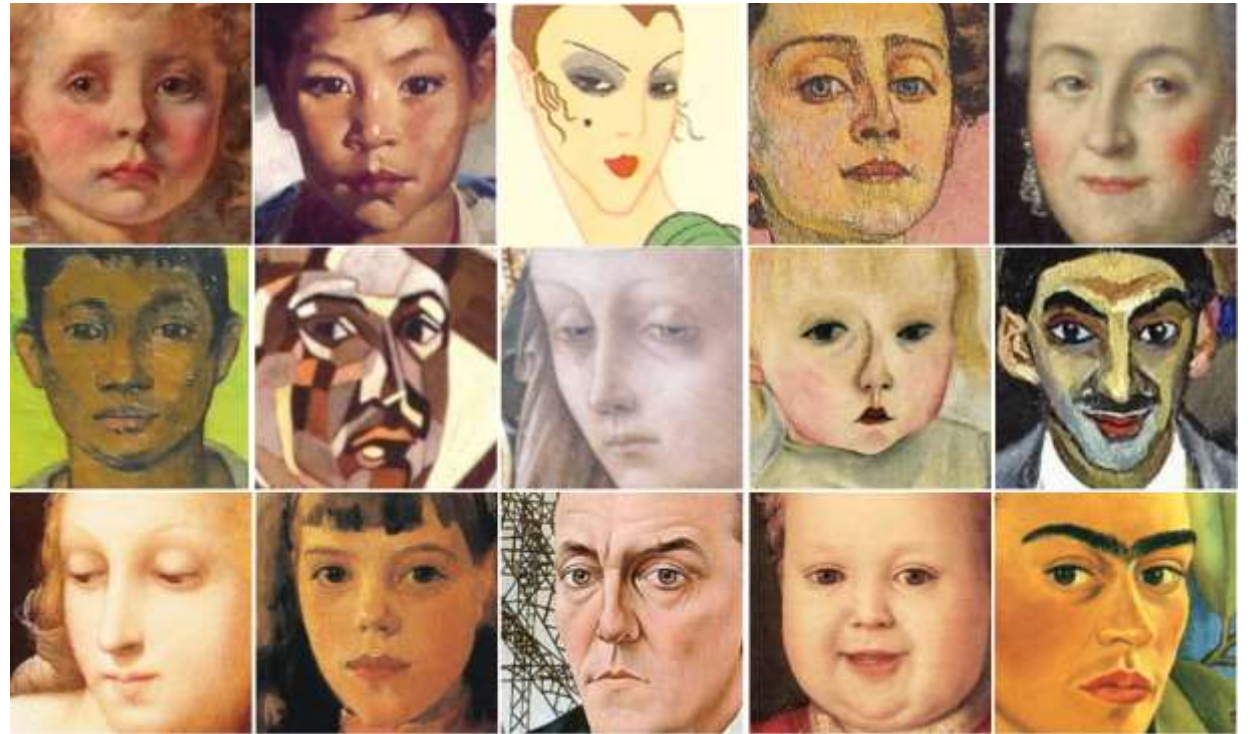


Datensatz

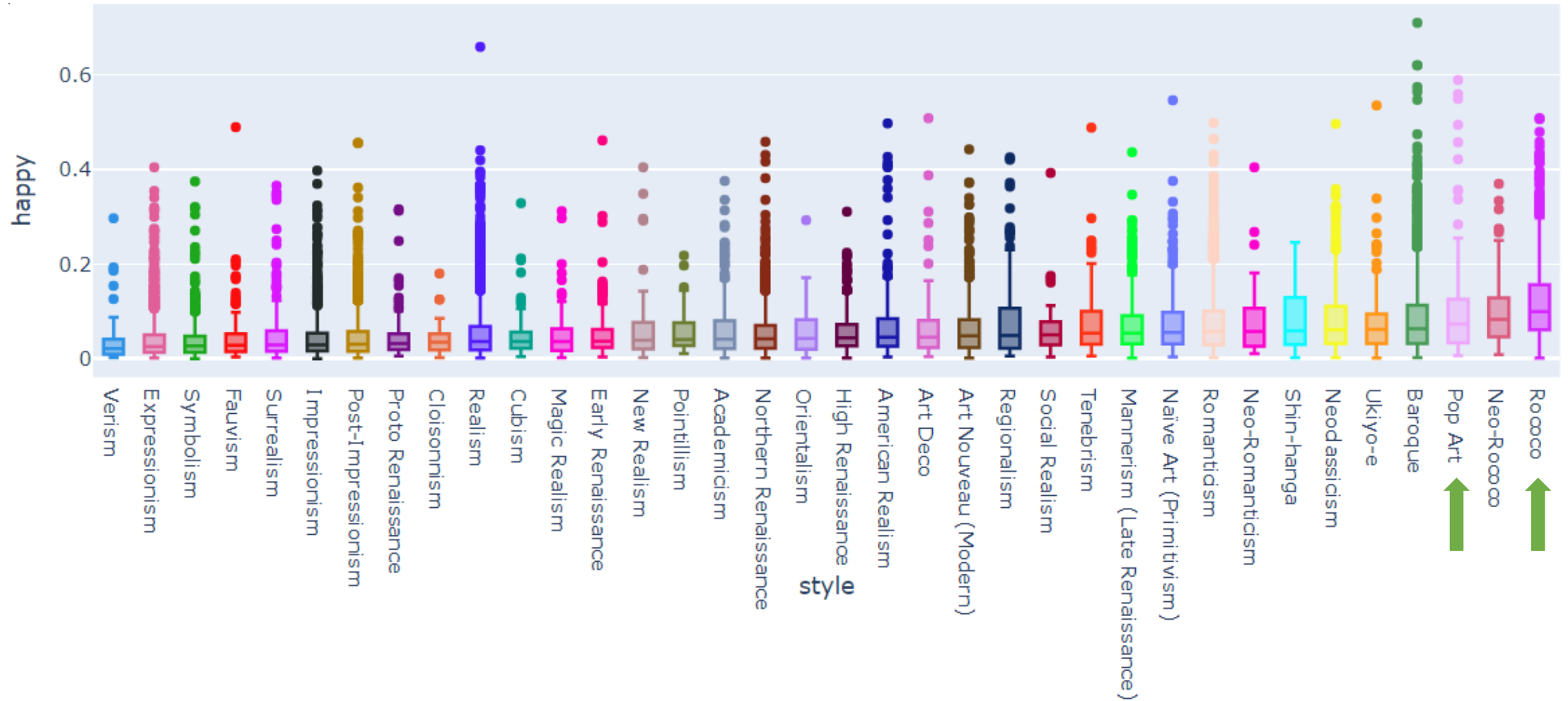
Erstellen eines Datensatzes auf Basis von Kunstwerken. Gesichter wurden extrahiert

- Portraits stammen von Wikiart.com
- Gesichtserkennung mit Single Shot Multibox Detector (SSD)

→ Ergebnis: 36 Klassen mit insgesamt >20.000 Gesichtsbildern



Statistische Auswertung



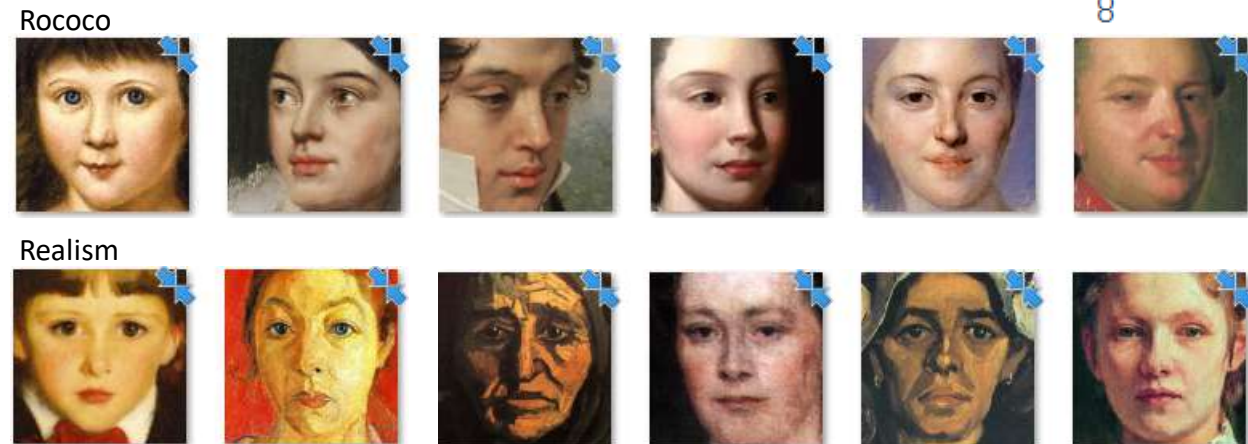
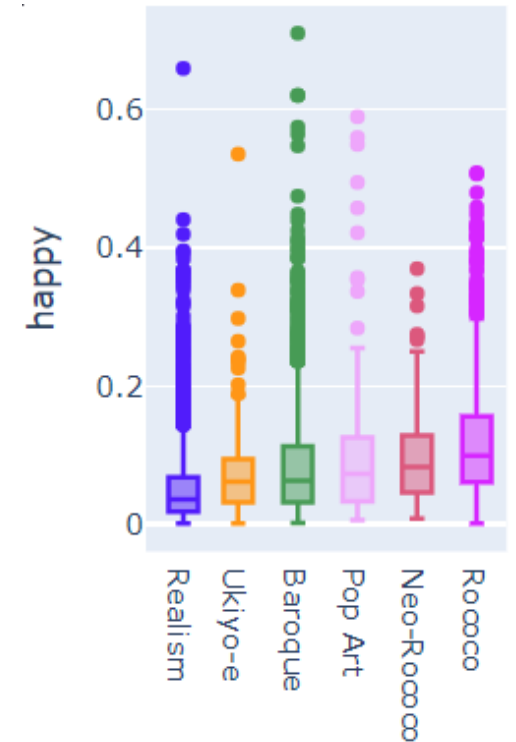
Statistische Auswertung

Statistische Auswertung von Labels und Geometrie (ANOVA und Games-Howell-Test)

Label ‚Happy‘:

→ Unterschied zwischen Rococo ($M=0.12$, $SD=0.08$) und Pop Art ($M=0.11$, $SD=0.11$)
signifikant? Nein: $p > .999$, (-0.01)

→ Unterschied zwischen Rococo und Realism ($M=0.05$, $SD=0.05$) signifikant? Ja: $p < .001$,
 $(0.07, 95\text{-CI}[0.06, 0.07])$



Erkenntnis: Die Darstellungen sind in Rococo
glücklicher als in Realism

Statistische Auswertung - Korrelationen

	date	sex	age	beard	makeup	pale	fatigue	angry	crying	surprised	mood	weight	beauty
date	1,00	0,10	0,02	0,07	-0,01	-0,15	0,00	-0,02	0,11	-0,03	-0,07	-0,03	0,03
sex	0,10	1,00	0,25	0,31	-0,13	-0,43	0,12	0,00	0,13	-0,30	-0,07	0,30	-0,49
age	0,02	0,25	1,00	0,37	-0,24	-0,28	-0,01	0,25	0,10	-0,24	-0,38	0,39	-0,56
beard	0,07	0,31	0,37	1,00	-0,21	-0,34	0,12	0,16	0,05	-0,25	-0,10	0,16	-0,35
makeup	-0,01	-0,13	-0,24	-0,21	1,00	-0,04	-0,06	-0,11	-0,04	0,13	0,20	-0,19	0,27
pale	-0,15	-0,43	-0,28	-0,34	-0,04	1,00	-0,07	-0,01	-0,18	0,25	0,02	-0,25	0,49
fatigue	0,00	0,12	-0,01	0,12	-0,06	-0,07	1,00	-0,02	-0,17	-0,62	-0,25	0,03	-0,13
angry	-0,02	0,00	0,25	0,16	-0,11	-0,01	-0,02	1,00	-0,07	-0,18	-0,24	0,20	-0,22
crying	0,11	0,13	0,10	0,05	-0,04	-0,18	-0,17	-0,07	1,00	-0,18	-0,11	-0,01	-0,11
surprised	-0,03	-0,30	-0,24	-0,25	0,13	0,25	-0,62	-0,18	-0,18	1,00	0,17	-0,09	0,34
mood	-0,07	-0,07	-0,38	-0,10	0,20	0,02	-0,25	-0,24	-0,11	0,17	1,00	-0,28	0,33
weight	-0,03	0,30	0,39	0,16	-0,19	-0,25	0,03	0,20	-0,01	-0,09	-0,28	1,00	-0,62
beauty	0,03	-0,49	-0,56	-0,35	0,27	0,49	-0,13	-0,22	-0,11	0,34	0,33	-0,62	1,00

Rococo

Expressionism

Korrelationen wie:

- beauty & age
- sex & age
- mood & age
- sex & pale

Column1	date	sex	age	beard	makeup	pale	fatigue	angry	crying	surprised	mood	weight	beauty
date	1,00	-0,01	0,19	-0,06	-0,12	-0,02	0,02	0,00	0,00	-0,13	-0,11	0,10	-0,01
sex	-0,01	1,00	0,26	0,40	-0,30	-0,29	-0,03	0,17	0,04	-0,10	-0,09	0,07	-0,39
age	0,19	0,26	1,00	0,29	-0,34	-0,19	-0,09	0,22	0,06	-0,17	-0,35	0,29	-0,40
beard	-0,06	0,40	0,29	1,00	-0,28	-0,19	-0,12	0,12	0,01	0,01	0,04	0,08	-0,16
makeup	-0,12	-0,30	-0,34	-0,28	1,00	0,01	0,07	-0,08	-0,03	0,00	0,12	-0,12	0,26
pale	-0,02	-0,29	-0,19	-0,19	0,01	1,00	-0,03	-0,19	-0,01	0,21	-0,05	0,03	0,32
fatigue	0,02	-0,03	-0,09	-0,12	0,07	-0,03	1,00	-0,29	-0,37	-0,63	0,06	0,04	-0,26
angry	0,00	0,17	0,22	0,12	-0,08	-0,19	-0,29	1,00	-0,04	-0,14	-0,02	0,10	-0,21
crying	0,00	0,04	0,06	0,01	-0,03	-0,01	-0,37	-0,04	1,00	0,12	-0,24	-0,08	0,04
surprised	-0,13	-0,10	-0,17	0,01	0,00	0,21	-0,63	-0,14	0,12	1,00	0,09	-0,07	0,38
mood	-0,11	-0,09	-0,35	0,04	0,12	-0,05	0,06	-0,02	-0,24	0,09	1,00	-0,22	0,12
weight	0,10	0,07	0,29	0,08	-0,12	0,03	0,04	0,10	-0,08	-0,07	-0,22	1,00	0,01
beauty	-0,01	-0,39	-0,40	-0,16	0,26	0,32	-0,26	-0,21	0,04	0,38	0,12	0,01	1,00

Statistische Auswertung - Korrelationen

Die Korrelationen sind teilweise sehr eindeutig und geben einen ersten Eindruck von Schönheitsidealen und Menschenbildern.

Beispiel:

Darstellung von weiblichen Personen meist: blass, jung, schön

Darstellung von männlichen Personen meist: alt, ernst

Darstellung von älteren Personen mit weniger positiven Emotionen

-> Dies zeigt sich in fast allen Kunststilen

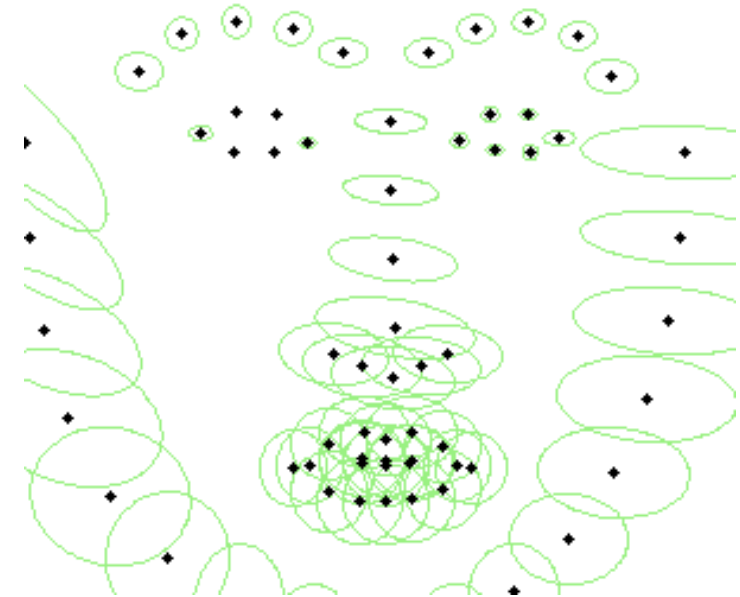
Geometrie der Gesichter

Semantische Beschreibung des Datensatzes durch Vergleich der Facial Landmarks und Betrachtung der Geometrie.

Beispiele (statistisch signifikant):

- In Ukiyo-e und Proto Renaissance sind die Nasen eher lang, in American Realism, Cloionnism und New Realism eher kurz
- In Regionalism ist der Mund am breitesten, in Early Renaissance am schmalsten
- Proto-Renaissance, Shin-Hanga und Ukiyo-e haben den größten Mund-Augen-Abstand
- Portraits in Art Deco haben eher ein schmales Kinn (kleiner Kiefer Winkel)

Proto Renaissance



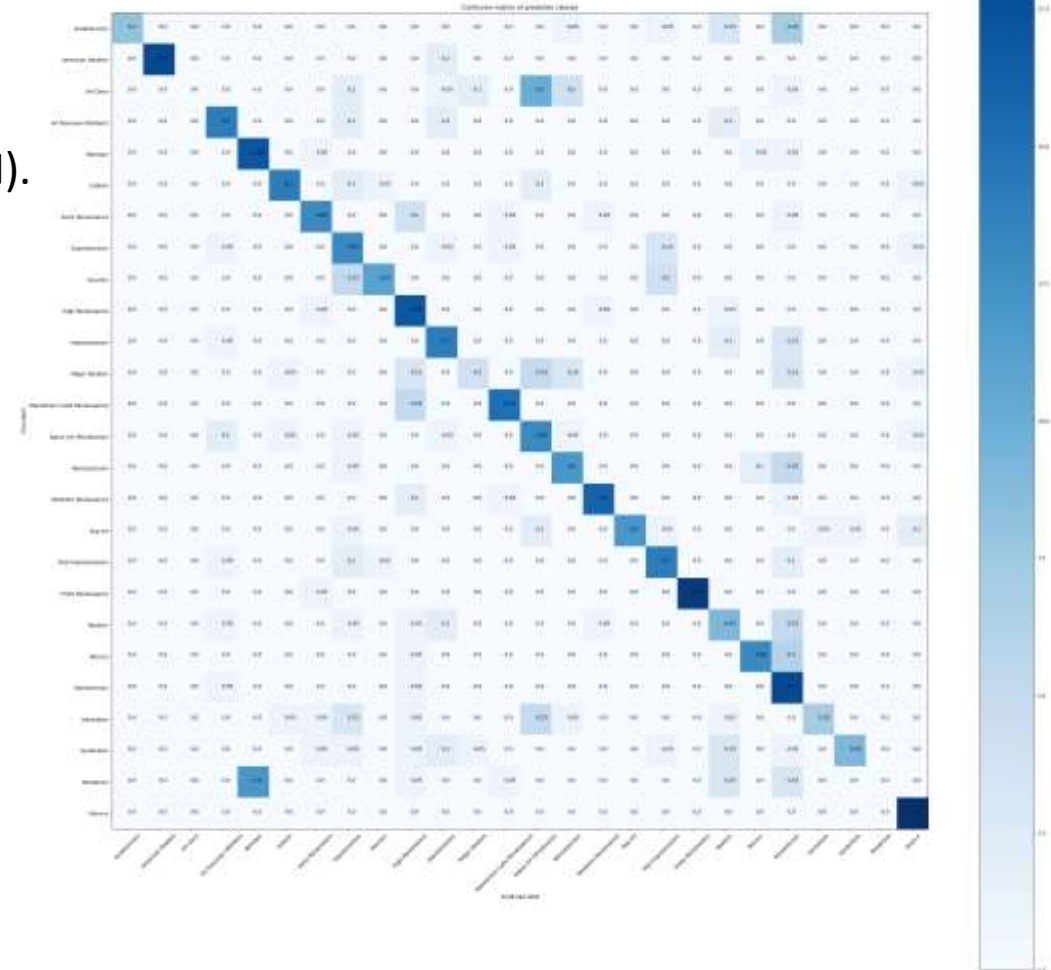
Klassifizierung

Klassifizierung mithilfe eines Convolutional Neural Network (CNN).
Verwendet wurde Resnet101V2, vortrainiert mit Imagenet-Daten

26 Klassen mit je > 100 Bildern:

- Training-Daten: 13665 + 5676 (validate) = 19.341
- Test-Daten: 20 je Klasse = 520

Ergebnis: training accuracy ca 68%
validation accuracy ca 57%
test accuracy ca 62%



Dimensionsreduktion mit t-SNE

Ansatz zur Dimensionsreduktion:

Die Features aus dem letzten Layer des Netzwerks mit 512 Dimensionen können mithilfe von t-SNE (*t-distributed stochastic neighbor embedding*) auf 2 oder 3 reduziert werden und visuell dargestellt werden.

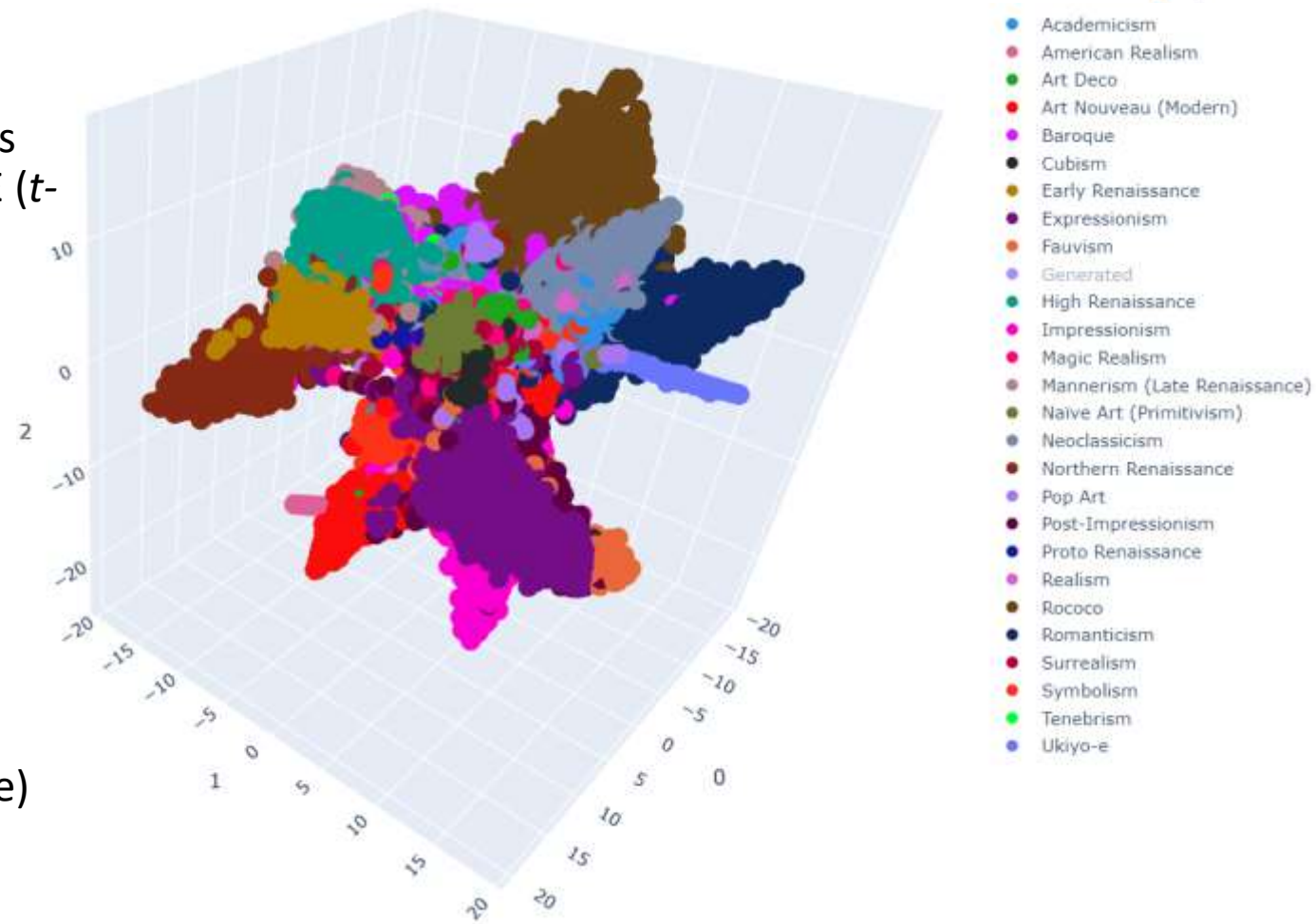
Beobachtung:

Es bilden sich deutliche Cluster für die einzelnen Klassen (siehe Bild)

Beobachtung:

Einige Klassen sind besonders distinkt (z.B. Ukiyo-e)

Andere gehen in einander über (z.B. Rococo und Romantizismus)



Dimensionsreduktion mit t-SNE

Beobachtung:

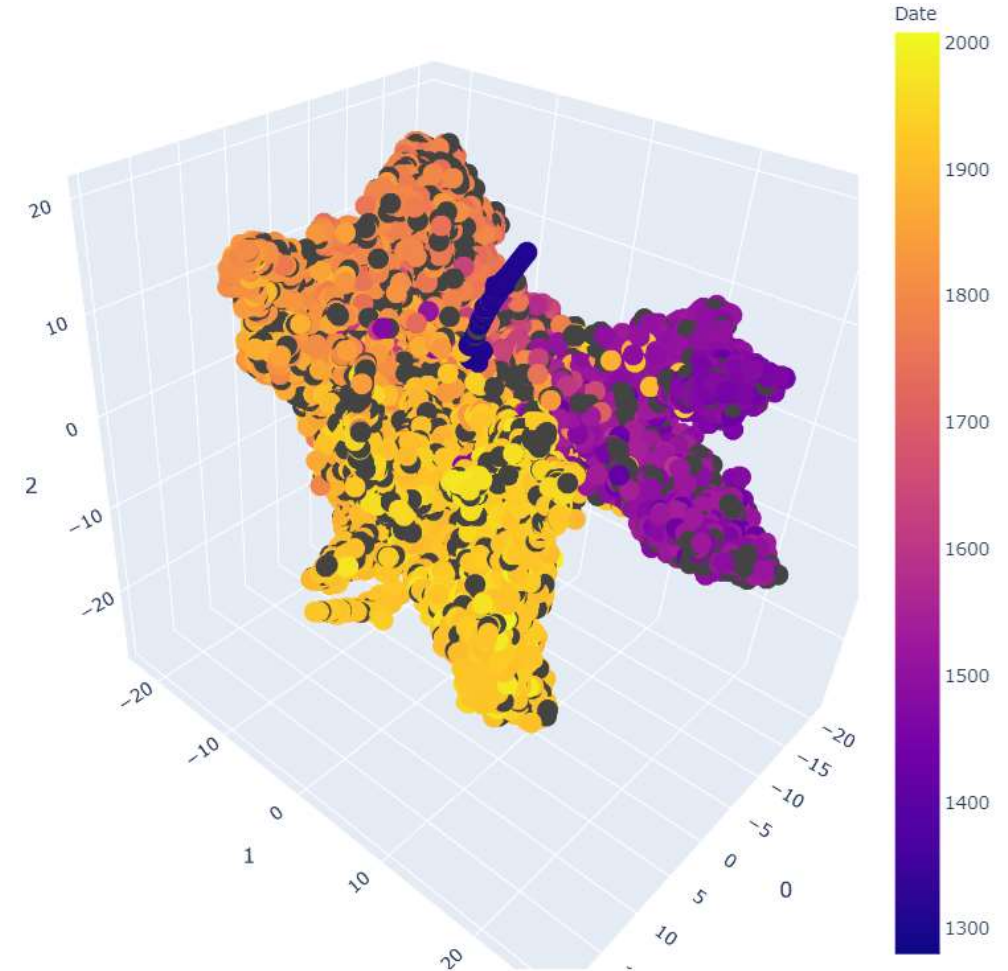
In der t-SNE-Verteilung werden bereits zeitliche Zusammenhänge sichtbar (erkennbar durch die räumliche Nähe von Datenpunkten derselben Farbe).

Daraus lässt sich eine fließende Veränderung von Merkmalen über die Zeit erkennen.

Lineare Regression (MLR) bestätigt dies:

$R^2 = 0.788$, $F(16, 10699) = 2496$, $p < .001$.

Die Features des Klassifikators können 78.8% der Varianz in der Datumsvariable erklären und der Korrelationskoeffizient $R = 0.89$ indiziert einen starken Zusammenhang.



schwarze Punkte: Kein Datum vorhanden

Methodik - Generierung



Generierung

Generierung von Gesichtern:

Generative Adversarial Networks (GANs) [4] sind Deep-Learning Modelle zum unüberwachten Generieren von realistischen Daten.

-> Training des StyleGAN2 [7], das bereits auf Gesichter vortrainiert ist

Ergebnis:

Die Portraits wirken sehr realistisch und divers. Nur wenige weisen Fehler auf

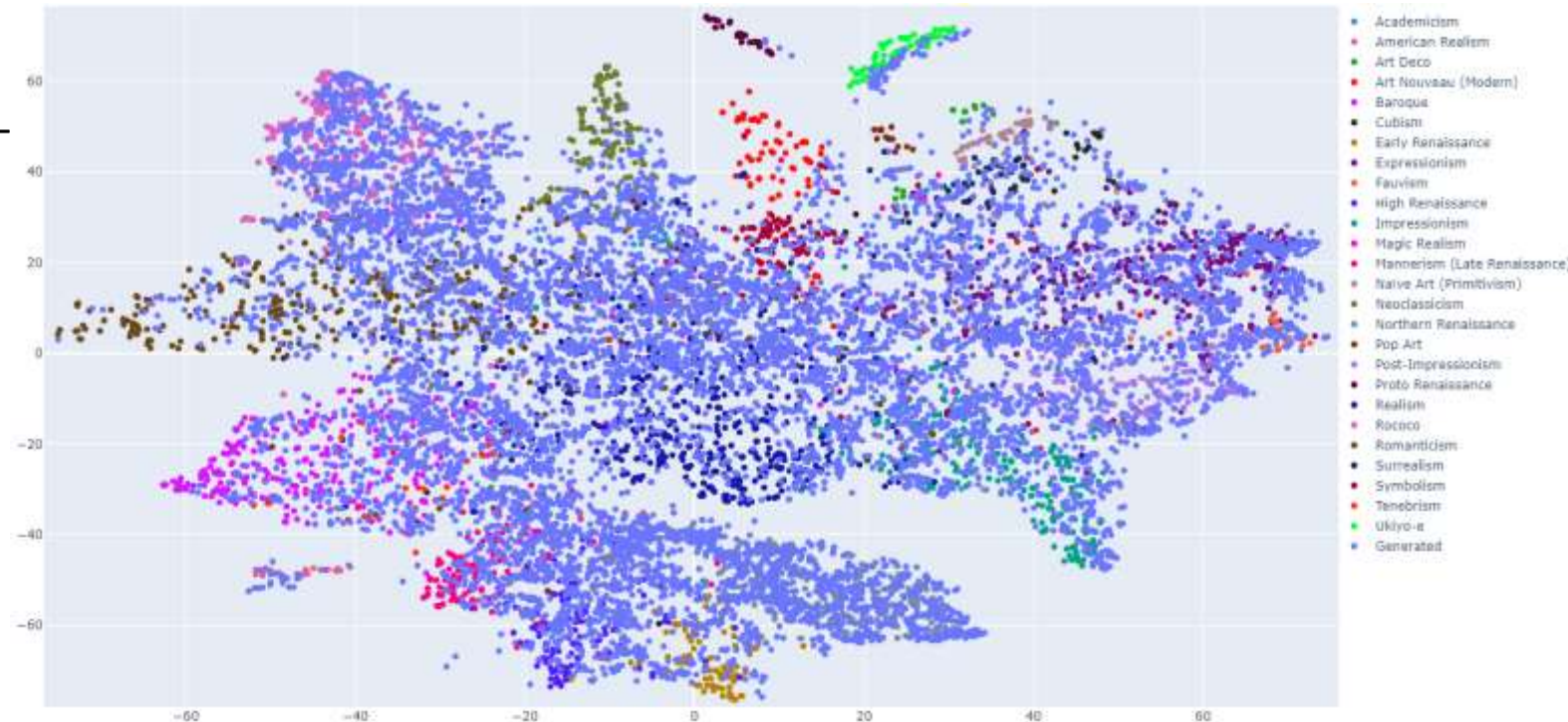


Generierung

Die generierten Daten werden ebenfalls mit dem Klassifikator analysiert und die Datenpunkte in die t-SNE Verteilung eingebettet.

Beobachtung:

Die ursprüngliche Verteilung wird nachgebildet, jedoch nicht perfekt rekonstruiert. Daten am Rand der Verteilung sind weniger vertreten, in der Mitte sind die Punkte dagegen sehr dicht.



Generierung

Beobachtung:

Die Verteilung wirkt teilweise verschoben. Das zeigt, dass das GAN sehr distinkte Merkmale zwar annähern, aber nicht komplett abbilden kann. Erkennbar z.B. bei den sehr distinkten Klassen Ukiyo-e und Naive Art:



-> diese Abweichungen könnten bereits als „neue Kunststile“ betrachtet werden, da sie den Stil einer Klasse nicht exakt treffen

Generierung

Beobachtung:

Bilder mit einer sehr geringen Vorhersagewahrscheinlichkeit passen auch nicht ganz in die Verteilung, sind aber eher mittig in der t-SNE Distribution angesiedelt und befinden sich somit zwischen verschiedenen Klassen



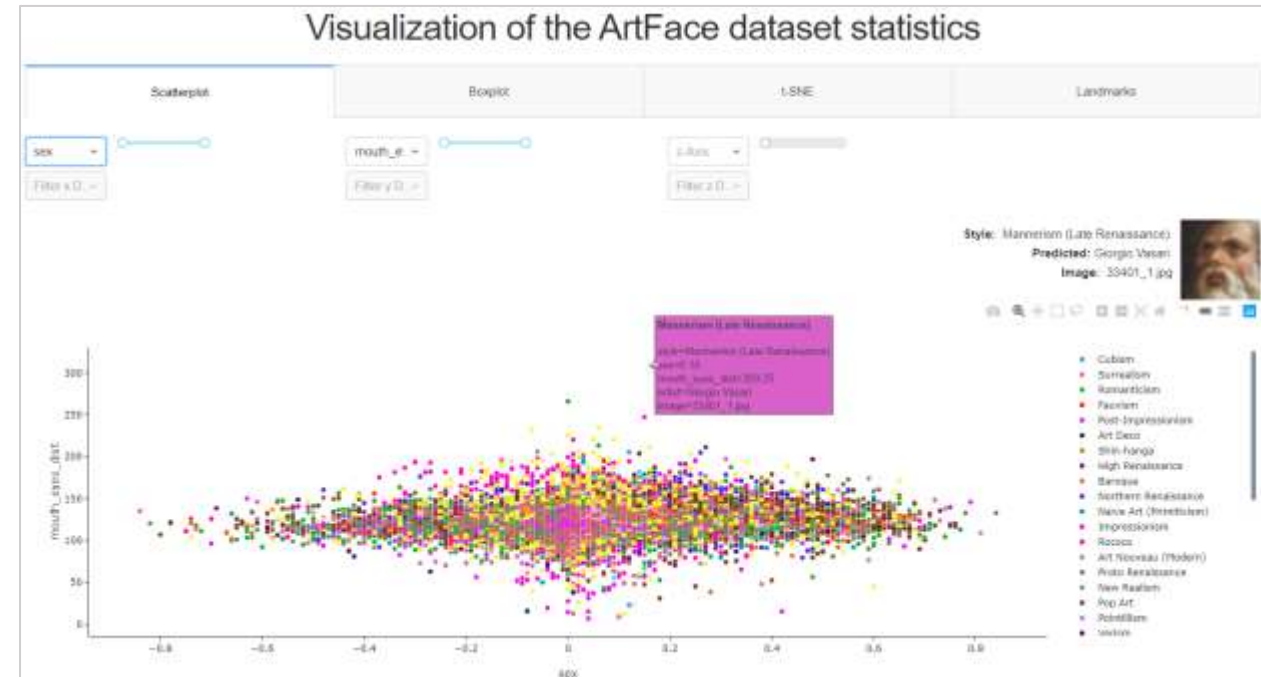
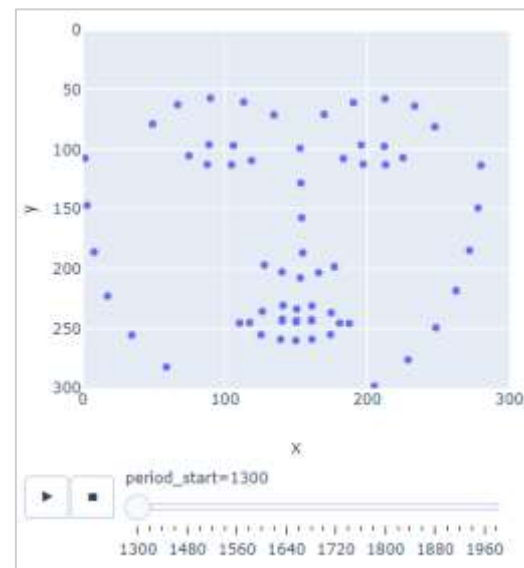
-> sie weisen sehr unspezifische Merkmale auf und vereinen Attribute aus verschiedenen Klassen (ebenfalls „neue“ Kunststile?)

Visualisierung

Visuelle Aufbereitung des Datensatzes und der Werte und Statistiken.

→ Interaktiv und flexibel

→ Ermöglicht einen besseren Einblick, tieferes Verständnis und macht Zusammenhänge sichtbar



Fazit

Deep-Learning Methoden können zur systematischen Analyse von kreativen Werken herangezogen werden. Sie bieten Betrachtungsweisen großer Datenmengen, die mit herkömmlichen Mitteln nicht möglich wären. In dieser Arbeit konnten unter anderem zeitliche Zusammenhänge und Schönheitsideale herausgestellt werden.

Vertiefende Untersuchungen sind jedoch nötig, um systematische Analysemethoden zu etablieren die auch mit Kunstwissenschaftlichem Wissen einhergehen.

GAN Modelle sind dazu in der Lage, realistische Portraits spezifischer Kunststile abzubilden aber gleichzeitig auch globale Interpolationen vorzunehmen und so augenscheinlich neue Stile zu erzeugen.



Literatur

- 1) Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. 2020. StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8188–8197.
 - 2) Frégier, Y. and Gouray, J.-B. 2019. Mind2Mind: transfer learning for GANs. *arXiv preprint arXiv:1906.11613*.
 - 3) Gatys, L. A., Ecker, A. S., and Bethge, M. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.
 - 4) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.
 - 5) Jenicek, T. and Chum, O. 2019. Linking Art through Human Poses. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 1338–1345.
 - 6) Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., and Winnemoeller, H. 2013. Recognizing image style. *arXiv preprint arXiv:1311.3715*.
 - 7) Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. 2020. Analyzing and improving the image quality of StyleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8110–8119.
 - 8) Lecoutre, A., Negrevergne, B., and Yger, F. 2017. Recognizing art style automatically in painting with deep learning. In *Asian conference on machine learning*, 327–342.
 - 9) Tan, W. R., Chan, C. S., Aguirre, H. E., and Tanaka, K. 2016. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE international conference on image processing (ICIP)*, 3703–3707.
 - 10) Tan, W. R., Chan, C. S., Aguirre, H. E., and Tanaka, K. 2018. Improved ArtGAN for conditional synthesis of natural image and artwork. *IEEE Transactions on Image Processing* 28, 1, 394–409.
 - 11) Taphorn, A. 2020. *Kunststilklassifizierung mit Convolutional Neural Networks*, Reutlingen.
 - 12) Yaniv, J., Newman, Y., and Shamir, A. 2019. The face of art: landmark detection and geometric style in portraits. *ACM Transactions on Graphics (TOG)* 38, 4, 1–15.
- a) CLIP: Connecting Text and Images, <https://openai.com/blog/clip/>