HON2200 Prosjekt 2

Nevral stiloverføring

Av Cassandra, Jonas, Halvor og Agnes

1 Introduksjon

Stil og innhold er de to aspektene som beskriver uttrykket til et bilde. Stil sier noe om farge, strek, tekstur, materiale og lignende, mens innholdet handler om objektene bildet viser. I 2015 ble artikkelen "A Neural Algorithm of Artistic Style" publisert, der Leon A. Gatys, Alexander S. Ecker og Matthias Bethge presenterer en metode for nevral stiloverføring. De har laget en algoritme som gjør det mulig å danne en representasjon av henholdsvis innhold og stil fra et bilde. Ved hjelp av denne teknologien kan man nå overføre stilen fra et bilde over på et annet. Denne algoritmen kommer til å bli tema for denne rapporten, der vi ønsker å undersøke skillet mellom innhold og stil fra tre perspektiver: begrepsbruk, algoritmisk og nevrologisk.

Vi kommer først til å legge frem teori som gir en redegjørelse for det konseptuelle synet på stil og innhold. Deretter kommer vi til å ta for oss det algoritmiske perspektivet. Her kommer vi til å ta i bruk metoden for nevral stiloverføring som ble introdusert av Gatys, Ecker og Bethge i 2015. Vi kommer til å forklare hvordan metoden benytter nevrale nett for å danne en representasjon av bilde og stil. Deretter vil vi legge vekt på hva algoritmen fanger opp som stil, ved å analysere resultatene fra utforskende arbeid med nevral stiloverføring. I diskusjonsdelen kommer vi til å koble resultatene fra vår egen kode til det konseptuelle perspektivet fra teoridelen. Er det vi ser på som stil det samme som det algoritmen fanger opp som stil? Og hva blir så denne algoritmens bruksområde med tanke på hva slags konsept for stil den fanger?

Avslutningsvis vil vi diskutere det nevrologiske perspektivet på stil og innhold. Gatys, Ecker og Bethge foreslår i artikkelen sin at denne metoden for nevral stiloverføring kanskje kan avsløre hvordan mennesket forstår estetiske uttrykk, ettersom metoden med konvolusjon ser ut til å gi resultater som passer med det vi tenker på som stil og innhold (referanse). Det antydes også at denne algoritmen har direkte overførbarhet til funksjoner i hjernen. Er det slik at det kunstige nevrale nettverket kan gi oss kunnskap om hvordan det våre biologiske nevrale nettet funker?

2 Teori

2.1 Begreper

I nevnte artikkel av Gatys, Ecker og Bethge, brukes begrepene "content" og "style" for å skille mellom hva algoritmen henter ut av informasjon fra hvert av bildene. Disse er begreper som gir intuitivt mening når vi ser på resultatbildene. Algoritmen har da vitterlig overført Munchs malestil i "Skrik" til et foto av hus. Artikkelforfatterne mener også selv at dette samsvarer godt med vår vanlige forståelse av begrepene når vi ser på et bilde. Likevel mener vi dette er begreper som bør være gjenstand for en (mer) finmasket undersøkelse for å fange opp ordenes nyanser. Selv om konsept og algoritme virker å overlappe, betyr ikke det at de er fullstendig sammenfallende.

For å redegjøre for hvert begrep, vil vi ta utgangspunkt i norske oversettelser av begrepene brukt av Gatys, Ecker og Bethge fordi dette ikke utgjør noe særlig oversettelsesproblem. Redegjørelsen for begrepene "style" og "content" i den amerikanske ordboken Merriam-Webster, sammenfaller i relevante betydninger med de norske "stil", "innhold" og "form" slik de er definert av Det Norske Akademis Ordbok (NAOB) (Merriam-Webster, s.v. "content", s.v. "form", s.v. "style", lest 12. mai 2022). Vår håper dermed vår diskusjon av de norske begrepene også vil være brukbar i engelsk sammenheng.

2.1.1 Stil

I følge NAOB er "stil" i kunst-, arkitektur- og musikksammenheng "form, formspråk, uttrykksmåte som er typisk for en periode, en retning (eller en enkelt kunstner, utøver)" (NAOB, s.v. "stil", lest 12. mai 2022). Her er definisjonen todelt ved både å karakterisere stilens innhold ("form, formspråk, uttrykksmåte") og attributt ("typisk for en periode, en retning (eller en enkelt kunstner, utøver)"). Form, formspråk og uttrykksmåte er noe vage termer som må utdypes for å forstås. I sin enkleste forstand er stil «måten noe gjøres på». Vi har et «noe» og dette fremføres, kommer til syne eller realiseres på en *måte.* Hvis vi også tar innover oss ordets betydning i annen bruk (som jo sier noe om nyanser og assosiasjoner til betydningen i kunstsammenheng), kan dette «noe» både være en handling («hun hilste i kjent stil») eller en livsførsel (livs*stil*) eller utformingen av estetiske uttrykk for eksempel i bilder, musikk, arkitektur eller litteratur.

Når vi trekker denne betydningen inn i bildekreasjon, kan måten innholdet fremstilles på, både tenkes å være knyttet til (a) teknikk for fremstilling - med andre ord det som knyttes til bildets materialitet (det ikke-representative, som ikke *nødvendigvis* sier noe utover seg selv);

men det kan også tenkes at fremstillingsmåte kan inneholde de (b) virkemidler som benyttes for å frembringe assosiasjoner, følelser eller intensivere innholdets kraft. Dette er kun to akser ved stilbegrepet, men som gjelder de samme komponentene i bildet. De to aksene overlapper i fysisk realisering, for eksempel penselstrøk, farger, detaljrikdom, nøyaktighet, materialtype (lerret, stein, oljemaling…), komposisjon, dynamikk (*claireobscure*-malerier for eksempel); men forskjellen er i hvilken grad de bidrar inn i verkets representasjon, innhold eller *mening.* For eksempel kan dottete, unøyaktige malingsstrøk være både knyttet til at dette er malerens foretrukne maleteknikk fordi den for eksempel er praktisk for henne (hvis motivet er bevegelig), eller det kan bidra til en følelse hos tilskueren av at dette dreier seg om en erindring eller et umiddelbart inntrykk, slik det er i de impresjonistiske maleriene.

Likevel, stil-begrepet gir uttrykk for noe mer enn bare «måte». Som vi ser av NAOBs definisjon henviser stil til en gjentakende samling trekk som er felles for flere uttrykk eller handlinger, gjerne av samme person eller tilhørende én retning eller gruppe mennesker. Slik er å benevne noes stil, å påkalle en skygge av kontekst og identitet til verket. Det ligger en slags tilleggsopplysning om at stilen er opphaverens måte å gjøre det på, og derfor kan settes i sammenheng med resten av det som deler samme stil. Det følger en tanke om at "hvis dette verket deler noe ved et annet verk, deler den også noe av det samme opphavet til det". Ikke nødvendigvis fordi det regnes som å være malt av samme person, men at det er noe som har fått to malere til å gjøre samme stilvalg bevisst eller ubevisst. Romantiske trekk i et bilde fra middelalderen gjør at vi tenker "begge malere har noe stilmessig felles, og er verdt å se i lys av hverandre". For eksempel i bildekunsten ville en «slik-og-slik–stil» si noe om hvem som har skapt bildet, eller hvilken tradisjon eller stilepoke den knyttes til. Å analysere stilen kan derfor være et verktøy for å vise hvordan bildet er typisk for den retningen eller opphaveren man allerede tilskriver verket, eller det kan være en måte å nettopp tilskrive det en likhet med en ellers urelatert kunstner, eller, tredje mulighet, det kan sette kunstneren alene som den første av sitt slag, sin stil. Uansett setter det bildet i en større sammenheng.

Men i en slik sammenheng, der stilen peker hen mot opphavet på et vis, vil flere kjennetegn enn bare de rene mikrostrukturene som reflekterer hvordan bildet er skapt på. Stil kan også være perspektiv, utsnitt (nært eller fjernt fra objektet), abstraksjonsnivå (er bildet gjenkjennelig som representasjon av noe i det hele tatt?) osv. Dette er virkemidler som igjen inneholder en god del mening og representasjonelt innhold *i seg selv.* Et nærbilde av et øye vil for eksempel tilsi at maleren har betraktet objektet sitt nøye. Dette forteller igjen noe om malerens interesse for objektet osv… Om dette kan regnes som stilkomponenter kan

diskuteres - ville man for eksempel si at en maler med en serie nærbilder, har en stil preget av inspeksjon? At det i det hele tatt er åpent for diskusjon, medfører at stilbegrepet også er blandet med representasjonelt innhold hvert fall.

2.1.2 Innhold

Allerede i ordet "innhold" kan vi skjønne hva det betegner til en viss grad. "Det som holdes i noe". Stort sett er det innholdet vi finner viktigst, for eksempel ved en brødboks, men ikke nødvendigvis. En brødboksfabrikk ville interessere seg mer for design og funksjonalitet ved boksen. Likevel, beholderen er der for å presentere innholdet; beholderen er der for innholdet. Det er altså en asymmetri i forholdet mellom innhold og beholder i deres *funksjon.* Vi finner en undergruppe av denne innhold - beholder-relasjonen ordboka. NAOBs definisjon nummer 3.1 av "innhold" er mer spesifikt til kunst:

> emne, stoff som behandles, uttrykkes i en (skriftlig eller muntlig) fremstilling, i en tanke, i et kunstnerisk verk ; handling(sgang) i en fortelling, en film e.l. ; informasjon, budskap i en ytring (i vid forstand) | til forskjell fra form" (NAOB, v.s. "innhold", lest 12. mai 2022).

Her er innhold og beholder vanskeligere å skille, selv om vi finner igjen den samme funksjonalitetsasymmetrien som ved brødboksen. Vi har et emne som behandles, og en fremstilling av innholdet. Kunstverket skiller seg fra brødboksen i at det ikke er to ulike fysiske komponenter, men ett fysisk element med et symbolsk innhold. Symbolsk i betydningen "står for noe annet enn seg selv". Det finnes mange analogier og termer for en slik symbolverdi - kall det "tegn", "mimesis", "bilde", "representasjon". Poenget er at stoffet - det fysiske uttrykket kan forstås som noe annet enn selve materialet det består av. En "o" er ikke bare en runding, men en lyd. Slik er rundingformen en beholder for lyden og bokstaven. Mens for en fugl som observerer et ark med rundingformer, vil ikke oppfatte det som noe mer enn sin egen materialitet - det har ingen *mening.* Slik er det det bildet etterlikner, det vi oppfatter som "objektet", som for oss får en slik mening, og som er innholdet i bildet.

Det er selvfølgelig nyanseringer av dette. For det første har vi bilder som ikke representerer noe - formalistisk kunst. Har ikke disse noe innhold? Hvis vi skulle beskrive Piet Mondrians "Composition A" fra 1923 ville det være et tolkningsspørsmål om firkanter og farger er innholdet, eller om det ikke har innhold. Igjen viser det begrepenes unøyaktige bruk. Stort sett er det likevel det objektet som representeres som er innholdet i bildet. Et annet tvilstilfelle er

hvor langt inn i tolkningen "innholdet" strekker seg. Innholdet i bildet er, i tillegg til å være for eksempel en person, også hva personen gjør, hvilken setting personen befinner seg i - kanskje skaper vi oss en historie om hva som skjer forut for øyeblikket i bildet? For et menneske som iakttar bildet, vil dette være en mulig beskrivelse av innholdet i bildet.

Vi står altså igjen med to begreper, stil og innhold, som begge har nyanser og utvaskede overganger. Vi kan kanskje heller snakke om en familielikhet mellom komponenter som passer inn i "stil" og komponenter som hører inn under "innhold". I hovedsak er det likevel det bildet representerer, som utgjør innholdet, og det ved bildet som ikke er representert av noe annet, men som selv representerer noe annet, som er stil.

2.2. Nevral stiloverføring

Nevral stiloverføring går ut på å bruke nevrale nettverk til å overføre stilen fra et bilde til et annet. Det gjøres ved å ta utgangspunkt i et "innholdsbilde" og et "stilbilde", der målet er å klare å ekstrahere innholdsinformasjonen fra det ene og stilinformasjonen fra det andre, slik at resultatet blir et bilde som er en kombinasjon av de to, altså et bilde som skal representere det samme som "innholdsbildet" men med stilen fra "stilbildet".

For å gjøre dette bruker vi et konvolusjonelt nevralt nettverk (forkortes til CNN på engelsk), som er et bildeklassifiseringsnettverk. I vår analyse bruker vi et ferdig trent konvolusjonelt nevralt nettverk kalt VGG19, den samme som brukt til å generere resultatene i artikkelen til Gatys, Ecker og Bethge (2015). Vi kommer derfor ikke til å trene opp et nytt nettverk til å gjøre stiloverføringen, men som vi snart skal gå nærmere inn på er det likevel noen parametere vi kan endre på for å påvirke resultatet av stiloverføringen.

Konvolusjonelle nevrale nettverk bruker filtre, som er matriser bestående av verdier som representerer forskjellige vekter, til å detektere ulike egenskaper i bildet, som linjer, kanter, og overganger mellom farger. Til sammen vil flere slike filtre kunne brukes til å detektere det vi betrakter som innholdet i bildet. Dette gjøres gjennom en rekke konvolusjoner, som går ut på at filteret flyttes over ulike deler av bildet og verdiene i filter-matrisen multipliseres elementvis med verdiene (pikselverdiene) i bildet i det området, for så å legges sammen til én verdi. Resultatet fra bruken av filteret er da en ny matrise, et såkalt "egenskaps-kart" ("feature map"), som inneholder verdiene fra konvolusjonene over hele bildet, som altså vil si noe om egenskapene nettverket har sett etter i bildet. Videre kan man stable flere slike filtre etter

hverandre for å finne mer komplekse egenskaper enn vi klarer med ett filter (ved å ta et nytt filter over utdata fra forrige konvolusjon).

Nettverket består av flere lag der samlinger av bildefiltre gir en utdata bestående av slike "egenskaps-kart", som da vil være filtrerte versjoner av bildet vi sender inn. I tillegg til filter-lagene består også nettverket av noen lag som blant annet fjerner støy, og noen lag som reduserer størrelsen på bildet. Gjennom disse prosessene vil vi bli kvitt mye informasjon i bildet vi sender inn, hovedsakelig relatert til stilen i bildet. Dette gjør at lag høyere opp i nettverket i mindre og mindre grad vil reprodusere pikselverdiene i det originale bildet, men i større grad fange opp det faktiske innholdet i bildet. For å få frem innholdet i bildet henter vi derfor ut informasjonen fra et av de siste/øverste lagene i nettverket.

Det å ekstrahere stilen fra et bilde er en litt annen prosess enn for innhold, siden det i dette tilfellet er nettopp de delene av bildet som vi for innhold ønsket å fjerne som vi ønsker å få frem. Stilrepresentasjonen oppnås ved å finne korrelasjonene i og imellom de ulike "egenskaps-kartene" fra de ulike lagene i nettverket. For et gitt "egenskaps-kart" finner vi korrelasjonene ved å beregne Gram matrisen, som er multiplikasjonen av matrisen med den transponerte av matrisen. Disse korrelasjonene vil kunne representere stilen i form av f.eks. hvilke farger som brukes og ulike teksturer. Videre vil vi rekonstruere stilen til bildet ved å ta hensyn til Gram matrisene fra de ulike lagene, der vi kan velge hvilke og hvor mange lag vi ønsker å ha med i stilrepresentasjonen. På samme måte som for innholdet vil kompleksiteten øke høyere oppover i nettverket. Jo flere lag vi velger å ta med, desto mer vil de lokale strukturene i bildet bli fanget opp i stil representasjonen, som vil føre til en jevnere, og i mange tilfeller bedre, representasjon. Vi vil altså kunne trekke ut ulike stilrepresentasjoner avhengig av hvilke lag vi velger å inkludere.

I prosessen med å generere det endelige bildet ønsker vi i størst mulig grad å minimere tapet av stil fra "stilbildet" og av innhold i "innholdsbildet", men for å klare å kombinere de to til ett bilde vil det måtte inngås kompromisser. Når stilen og innholdet skal kombineres velger vi derfor også i hvor stor grad hver av delene vektlegges, der ulike vekter vil kunne gi veldig ulike resultater. Hvilke vekter som gir det "best" bildet vil være en individuell vurdering og kan variere mye avhengig av hvilke bilder vi kombinerer og hva det er vi ønsker å få frem.

3. Metode

For å gjennomføre nevral stiloverføring bruker vi et *pythonbibliotek* kalt *pystiche*. For å spare beregningstid bruker vi som nevnt et ferdig opptrent konvolusjonelt nettverk kalt VGG19. Vi definerer en *loss-funksjon* som en sammensetning av tap av stil, og tap av innhold. Hvor mye hver av disse loss funksjonene skal vektes bestemmes av parametere. Vi lar den iterere et visst antall ganger hvor den vil jobbe for å minske den totale loss funksjonen, dette definerer vi som antall steg algoritmen tar, og vil påvirke beregningstiden. Den siste parameteren vi i oppgaven vil endre er antall stil-lag vi lar algoritmen jobbe med/på.

Vi velger å gjøre selve stiloverføringen til en samlet funksjon i *python*, hvor vi tar inn de parameterne nevnt ovenfor, sammen med en kontekst og et stilbilde, for å så returnert det stiloverførte bilde.

Pystiche-pakken inneholder demobilder og som en start vil vi prøve og mikse noen av disse sammen for å se nærmere på resultatet. Vi vil også utforske hvordan antall itereringer algoritmen får lov til å gjennomføre påvirker resultatet og tiden den bruker. Til dette bruker vi et bilde av en fugl som innholdsbilde og mosaikk som stilbilde.

Vi utforsker også noen bilder utenfor demo pakken som vi tror, sammen med visse stilbilder, vil endre følelsene bilde vekker. Dette får oss til å gå nærmere inn på hvordan stil faktisk er representert i de ulike lagene. For å få en bedre forståelse av hvilke deler av bildet som er stil, velger vi ut oss et knippe bilder som har det som vi ser på som ulik stil, hvor vi også har en oppfatning av at bildene inneholder stil på ulike måter. Vi genererer et støybilde som vi bruker som innholdsbilde, deretter overfører stilen til hver av stillbildene og setter vekten til innhold til null. Dette lar oss undersøke hva de ulike lagene faktisk henter ut av bildene.

Vi lager til slutt en matrise for å undersøke hvordan vektlegging av stil og innholds-tap, og valg av lag, sammen påvirker resultatet av stiloverføringen. Kontekst bilde vi her bruker er hentet fra mappen med mat-bilder gitt i oppgaveteksten (Kilde?). Stil-bilder er hentet med tanke på høytiden eller bruksområdet til stiloverføringen vi velger. Dette vil muligens ha liten overføringsverdi til andre innhold og stil-bilder, men det vil gi oss en bedre forståelse av hvordan disse parameterne sammen påvirker resultatet.

4. Resultater

Som forventet ble resultatet av stiloverføringen bedre dersom vi lot algoritmen få iterere flere ganger. Dette kan vi se på bildene under, hvor vi først lot den kun iterere 10 ganger (se figur 3) også 500 ganger (se figur 4). Siden algoritmen tar utgangspunkt i innholdsbildet vil stilen fra mosaikk bilde kommer tydeligere fram når vi lar den kjøre 500 ganger. Her har vi latt vektene og antall stil-lag være konstant. Figur 1 er innholdsbildet og figur 2 er stilbildet brukt til testingen.

Figur 1 Innholdsbilde

Figur 2 Stilbilde

Figur 3 Stiloverført bilde. Antall iterasjoner = 10

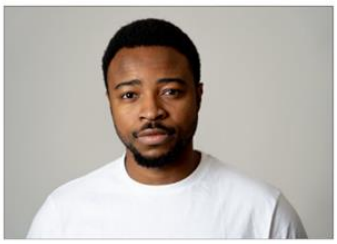
Figur 4 Stiloverført bilde. Antall iterasjoner = 500

Våre første erfaringer med notebooken ble gjort i henhold til oppgaven som ble gitt som forslag i oppgavebeskrivelsen, nemlig at vi ved hjelp av stiloverføring skulle tilpasse helt ordinære bilder av matprodukter til reklame for Oda til diverse høytider. Altså skulle det overføres stiler slik at matproduktene skulle selge bedre enten i jula, påsken, 17. mai eller lignende. Vi utforsket først med fuglebilde i mosaikkstil, som ikke ble så verst etter et par justeringer. Så prøvde vi å omgjøre ”Skapelsen av Adam” til kubistisk stil, denne var ikke særlig vellykket, kun noen minimale endringer i tekstur og farge. Deretter prøvde vi å lage noe som kunne funke som reklame for 17. mai: en kake i norgeflagg-stil. Denne ble helt ok,

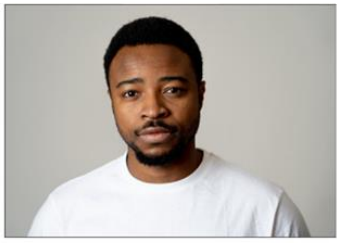
men vi så ikke noen særlig grunn til at denne skulle øke salget av kaker noe betraktelig. Etter to-tre ekstra forsøk på å stiloverføre landet vi på at denne oppgaven både var vanskelig å få til og ikke minst kjedelig.

Derfor endret vi retning og satte oss en ny problemstilling: er det mulig ved stiloverføring å vekke ulike følelser med det samme innholdsbilde? Vi tok utgangspunkt i et bilde av en mann med verdens mest nøytrale ansiktsuttrykk, for så å se om stiloverføringen kan vekke ulike følelser, men igjen uten hell! Stilbildet var dystert og med et dystert innhold, men det dystre ville ikke overføres mer enn i fargen. Bilde, som vist, ble bare i svart-hvitt. Da vi prøvde å gjøre den nøytrale mannen glad, ble han regelrett miserabel i det første forsøket, mens man heller ikke fikk noen merkbar endring i det emosjonelle utrykket for andre forsøk.

Figur 5 Stilloverføring(nederst) med innholdsbilde (øverst), stilbildet (i midten).

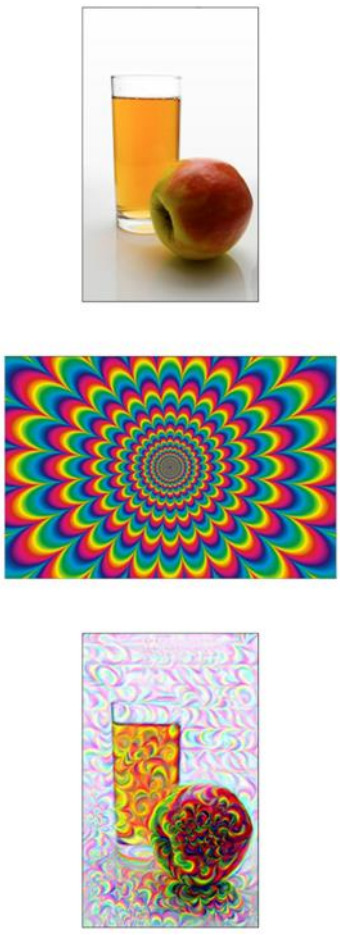


Figur 6 Stiloverføring(nederst) med innhold fra et nøytralt ansiktsuttrykk(øverst), med et negativt følelsesladd stilbilde (i midten).



Figur 7 Stiloverføring (nederst) av et positivt føllsladd stilbilde(midten) på et nøytralt ansikt innholdsbilde (øverst)

Altså var ingen av disse stiloverføringene særlig vellykket. Så da var spørsmålet: hva gjør vi nå? Vi må finne en stil som lar seg overføre, hvis ikke kommer vi ingen vei! Så kom heldigvis gjennombruddet. Vi fant en stil som virkelig lot seg overføre: fargespillet som kjennetegner en LSD-rus, slik som illustrert på bildet under. Vi prøvde å overføre det til et bilde av eplejuice og ble endelig fornøyde. Her hadde stilen hatt en betraktelig innvirkning på bilde, samtidig som innholdet var tydelig bevart. Dette ga oss en idé. Hva hvis vi lager en ny høytid som Oda skal lage reklame for. Vi kaller den bare for høy-tid, av åpenbare grunner.

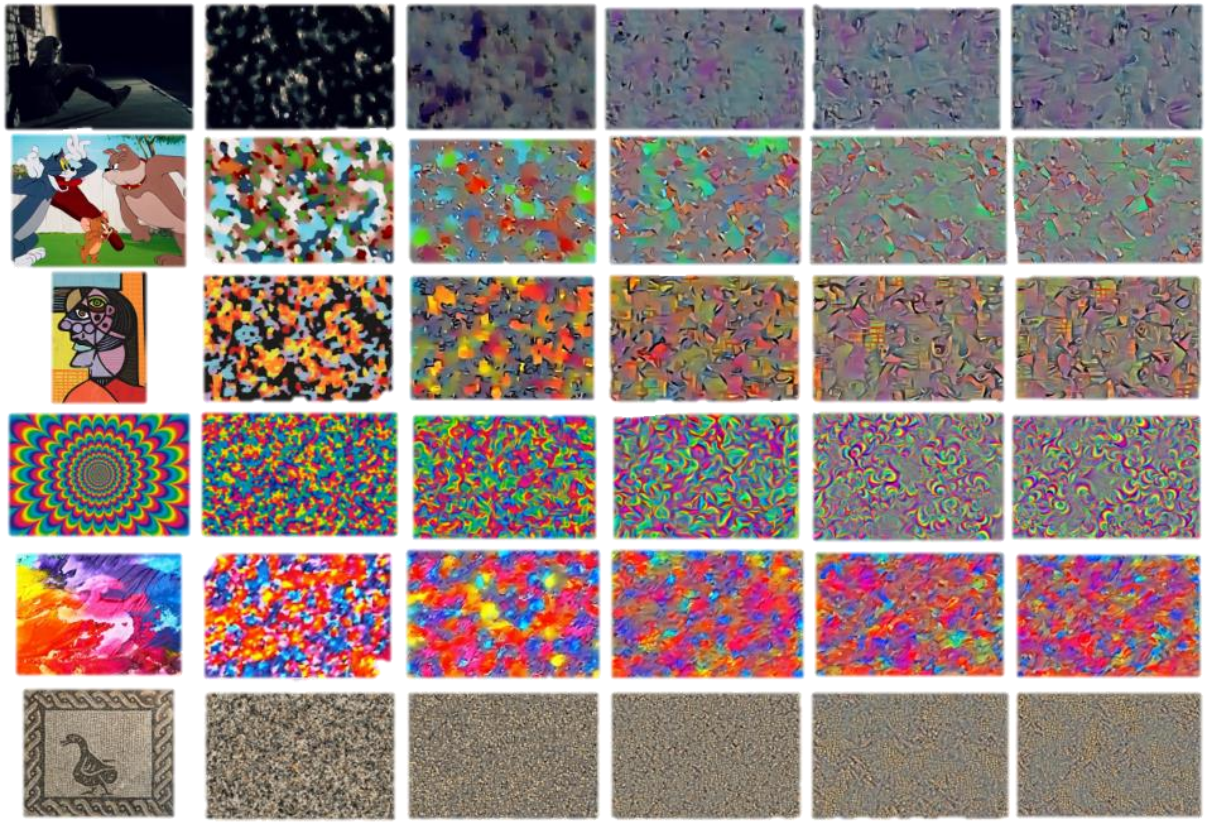


Figur 8 Stiloverføring (nederst) av «LSD-stil» (midten) på et innholdsbilde (øverst).

Vi endte til slutt med å skrinlegge denne idéen også, fordi vi innså at denne reisen vi hadde vært på hadde lært oss noe mer interessant. Det vi så på som stil før vi startet prosjektet er kanskje en misforståelse av begrepet. Og vi hadde i alle fall misforstått algoritmens konsept for stil og ikke minst hvilke bilder som egnet seg godt for stiloverføring. Denne åpenbaringen ledet oss til vår nåværende problemstilling. Hva er skillet mellom innhold og stil? Hvilken del av vårt stilbegrep er det algoritmen fanger opp? Og hva har dette å si for senere bruksområde?

For å undersøke hva algoritmen fanger opp kumulativt i konvolusjonslagene, lagde vi en visuell framstilling av stilrepresentasjonen fra de ulike lagene (figur 9). Man ser at algoritmen plukker ut ulike aspekter av ulike bilder ved ulike lag. I det første og laveste laget henter den i all hovedsak ut fargene og gradvis henter den ut mer av tekstur og mønster i bilde, for å danne en stilrepresentasjon. Denne utviklingen med større grad av mønstre er særlig tydelig for bildene med kubisme og LSD-trip. For Tom & Jerry-bildet blir det etter hvert innslag av små føtter, bein og øyne. Der har innholdet tatt noe overhånd. Det samme gjelder til en mindre grad for det kubistiske bildet. På det første bilde ser vi derimot at den ikke fanger opp stort annet en farger, det var dette som gjorde at vi ikke fikk noe godt resultat da vi prøvde å stiloverføre en trist emosjon tidligere. I bildene for maleri og mosaikk kommer det i større grad frem en representasjon av en teknikk,

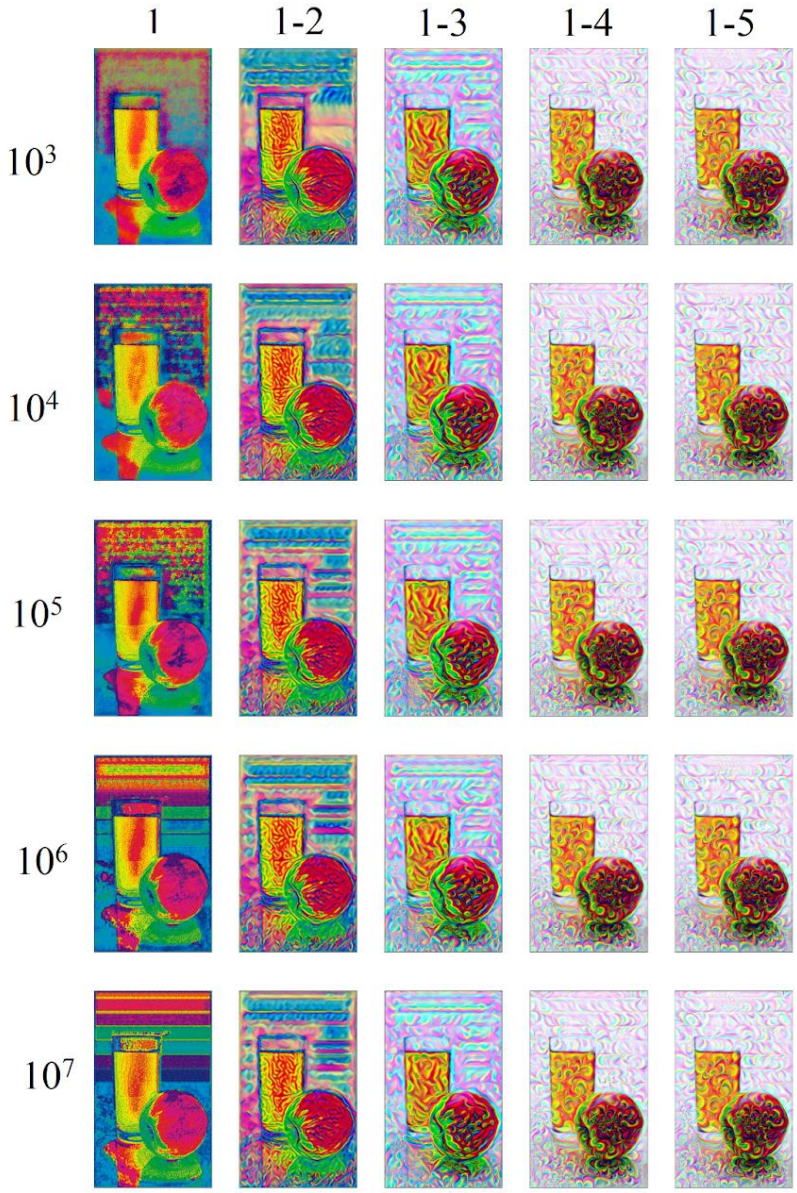
som er interessant. Man ser i større grad at det er antydninger til henholdsvis penselstrøk og mindre mosaikkfliser.



Figur 9 Stilbilder til venstre stiloverført på et svart-hvitt støybilde. Stiloverføringen er gjort med ulike «stillag». Tilvenstre nærmest original bildet, er det kun brukt de laveste stillagene, mens ett og ett lag er lagt til i stilanalysen for hvert bilde mot høyre.

Ut ifra disse bildene er det klart at metoden som er lagt frem av Gatys, Ecker og Bethge i større grad fanger opp mønstre og farger (og evt. teknikk). Vil man altså ha en vellykket stiloverføring er det best å velge stilbilder med tydelige mønstre og gjerne en annen fremstilling enn ordinære fotografier. Erfaring med algoritmen viser at den fungerer best med malerier som stilbilder. Det algoritmen ser på som stil fanger dermed en eller flere, men ikke alle, aspekter av det vi ser på som stil (som vist til i teoridelen). Den fanger opp farger, struktur, teknikk og mønster, særlig for malerier. Men den har noen mangler ved stil som en måte å fremstille noe på. For eksempel stil som å endre på størrelsesform eller formen på et objekt. Dette er særlig tydelig for tegneserier og i noen mindre grad for kubisme. I en tegneserie for eksempel får ikke algoritmen frem abstraksjonsprosessen av objektet. Den fungerer i det henseende dårlig som et 'filter', det kommer nok av at den ikke i like stor grad oppfatter hvilke objekter som er viktige i et bilde.

Til slutt, etter å ha undersøkt hvilke stiltyper som best fanges opp, og i hvilke lag de ulike stilkomponentene fanges opp i, undersøkte vi hva som ga best stiloverføring over på innhold. Vi valgte en stil som vi mente algoritmen fanget opp godt, og undersøkte hvordan den best skulle overføres på et annet bilde ved å endre på to parametere: 1) forholdet mellom *content loss* og *style loss*, og 2) hvilke lag vi tok med i stilanalysen. Resultatet ses i figur 10. Det er en tydelig tendens til at antall lag som tas med i stilanalysen har mer betydning på resultatet enn forholdet mellom stil og innhold, gitt de forholdstallene *vi* brukte. Likevel, høyere vekting av stil (de nederste radene), ga litt ulikt utslag avhengig av hvilke lag som ble regnet med. Ved færre stillag (venstre side av tabell), gir økt vekting av stil mest utslag på bakgrunnen i bildet, hvor det former seg vannrette linjer. Disse kommer fra det originale stilbildet og ikke innholdsbildet. Når vi beveger oss mot høyere lag, gir høyere stilvekt ikke like mye utslag, men vi ser at virvelmønsteret blir noe klarere.



Figur 10: Bildene i figuren viser resultatene etter ulike valg av forhold mellom vekter, og med ulike valg av lag brukt i stilrepresentasjonen. Verdiene til radene angir forholdet s/i, der s er vekten til stil og i vekten til innhold. Verdiene til

Videre gir bevegelsen fra å inkludere få og lave lag i stilanalysen til å inkludere høyere lag, en større estetisk forskjell. Her ser vi, i tråd med analysen av figur 9, at ved å inkludere høyere lag, får vi mer mønsteret fra den stilbildet, og ikke bare fargene. I valget mellom hva som gir best resultat, må bruksformål og estetisk smak spille inn. Vi synes i dette tilfellet at bildene med høy stilvekt og høyere stillag ga det mest effektfulle uttrykket.

4. Diskusjon

4.1 Hva sier analysene våre om bruksverdien av nevral stiloverføring?

Våre - kanskje lite kreative - forsøk på å gi innholdsbilder ulik stil etter humør eller sesong (17. mai og HØYtid) kan vi ikke konkludere med at har noen særlig bruksverdi. Dette skyldes at vi ikke kunne komme på noen stiler i betydningen "små- eller mellomstore mønsterstrukturer, annet enn farger og motiver (flaggene i figur5), som reflekterte følelse eller stemning.

Som vi fant i visualiseringen av forholdene mellom egenskapskartene (figur 9), er det mikrostrukturer og gjennomgående mønstre i tillegg til farge som fanges opp av algoritmen. Slik er bilder med en nærmere realistisk stil, eller en stil som ikke ses som gjennomgående mønster, ikke så godt egnet som stilbilder. Se bare på figur 6, hvor vi har et nesten fotorealistisk bilde som stilbilde i håp om å fange opp farger og komposisjon mellom lys og mørke. Det beste resultatet fikk vi av å bruke bildet med regnbuesirkler som stilbilde (figur 9 og 10). Her er det altså et bilde med kun form og farge, ikke noen representasjon - ikke noe særlig innhold. Vi synes dette fungerer bra som stilbilde, men samtidig mister stiloverføring noe av sin verdi dersom det som fungerer best er bilder med en sterk mikrostruktur. Vi må ta høyde for at vi ikke har funnet rett balanse mellom hvilke stillag vi inkluderer og vektingen av stil versus innhold. Likevel mener vi at med disse kravene til stilbilder som nevnt over begynner verdien av metoden å minke. Kunne vi få liknende resultater med en filterteknologi: mosaikk-filter, regnbue-filter eller impresjonisme-filter?

Verdien av den nevrale stiloverføringen er likevel til stede der det ikke bare er det estetiske resultatet i seg selv vi ønsker. Først har teknologien flere kunstneriske verdier. Vi har for eksempel et konseptuelt interessant aspekt ved at kunstneren er mer distansert fra prosessen av den estetiske utformingen. Dette gir et slags overraskelsesmoment ved resultatet, i tillegg til at kunst laget av "kunstig intelligens" åpner opp for tenkning rundt kunst og rollen mennesket har på seg selv som kunstner. Dernest har vi, i forlengelsen av disse konseptuelle

aspektene, et verdimoment i at man kan trekke ut stiler fra kjente verk. Hvis stilen som overføres er typisk for et bilde med kulturell eller symbolsk verdi, vil gjenkjennelsen av det ene bildet i det andre ha en verdi for betrakteren. Til sist har nevral stiloverføring en vitenskapelig verdi ved at den poengterer analysemetoder for bilder. Slik må vi kanskje revurdere eller nyansere våre egne oppfatninger av stil og innhold, samt at den kanskje sier oss noe om hvordan vi biologisk oppfatter kunst - hvis nevrale nett faktisk etteraper hjernens struktur.

4.2 Overførbarhet til biologi

I artikkelen til Gatys, Ecker og Bethge påstås det at algoritmen brukt i denne oppgaven kan brukes til å gi en forståelse av hvordan nevrale representasjoner kan fange forskjellen mellom innhold og stil. Altså at man, gjennom å studere hvordan algoritmen fungerer, kan få en forståelse av hvordan denne representasjonen blir til i hjernen også. Det antydes at nevronene i hjernen behandler bildene på en lignende måte som denne maskinlærings-algoritmen. Slik forklare de hvordan dette kan være mulig:

> In fact, our work offers an algorithmic understanding of how neural representations can independently capture the content of an image and the style in which it is presented. Importantly, the mathematical form of our style representations generates a clear, testable hypothesis about the representation of image appearance down to the single neuron level. The style representations simply compute the correlations between different types of neurons in the network. (Gatys, Ecker og Bethge 2015, 8-9)

Forfatterne påstår at ved å ha en algoritme som er basert på et nevralt nettverk (som er modellert etter hvordan nevronene fungerer i hjernen), kan man analysere denne algoritmen helt ned på nevron-nivå for å få et innblikk i hvordan slike nevrale nettverk fungerer. Spørsmålet er bare om denne analogien holder mål. Det er klart det er likheter mellom kunstige og biologiske nevrale nettverk. De kunstige nettverkene lærer av de biologiske stadig vekk, men kan kunnskapsutbyttet gå begge veier?

Antydningen til Gatys, Ecker og Bethge underbygges ved å referere til komplekse celler i det primære visuelle feltet V1, som ligger helt bakerst i hjernebarken. (Carandini 2012) I motsetning til enkle celler, kan disse komplekse cellene aktiveres ved et spekter av ulik stimuli. Enkle celler responderer kun på én form for stimuli. Videre trekker de inn “visual ventral stream” (her har vi ikke funnet noen god norsk oversettelse og vil utover kun referere til VVS):

> Extracting correlations between neurons is a biologically plausible computation that is, for example, implemented by so-called complex cells in the primary visual system (V1). Our results suggest that performing a complex-cell like computation at different processing stages along the ventral stream would be a possible way to obtain a content-independent representation of the appearance of a visual input. (Gatys, Ecker og Bethge 2015, 9)

Det er her analogien virkelig tar form. VVS går nemlig fra V1 inn til områder av hjernen som er ansvarlig for språk og hukommelse (Kravitz et al. 2013). Den kalles også for "what-stream", fordi dens hovedfunksjon er å identifisere objekter og finne ut hva ulike objekter er. Altså gjør den mye av det samme som algoritmen gjør når den bygger sin innholds-representasjon. Interessant nok påstår Gatys, Ecker og Bethke at man kan få en innholds-avhengig representasjon av den visuelle inputen ved å analysere ulike "lag" langs VVS. Altså insinueres det at denne strømmen fungerer lagvis på samme måte som algoritmen, i at nevronene fokuserer på ulike aspekter ved den visuelle inputen ved ulike lag. Ergo helt analogt med det kunstige nevrale nettverket.

Det er en spennende antydning, men grunnlaget for antydningen virker noe spekulativ. Kilden som er referert til av Gatys, Ecker og Bethge for å underbygge påstandene i denne delen av artikkelen handler kun om objekter i bevegelse og virker ikke spesielt relevant for kategoriene innhold og stil. Videre er det et nok usannsynlig å få testet hypotesen med det første, men prosjekter som Neuralink viser at det nok ikke er like langt fram i tid som noen av oss kanskje skulle ønske (Neuralink 2022).

5. Konklusjon

Vi har analysert forholdet mellom stil og innhold fra tre ulike perspektiver. Den begrepsmessige analysen viste at begrepet "stil" referer til måten noe fremstilles på, mens "innhold" referer til det som blir fremstilt. "Stil" er likevel en bred samling av trekk som er vanskelig å skille fra innhold i og med at det ikke bare referer til måten i seg selv, men er iboende knyttet til en person, gruppe eller kunstretning. Deler av denne brede samlingen av trekk fanges opp av algoritmen. Avhengig av hvilke lag vi bruker som utgangspunkt for stilanalysen, fanger den opp farger, materialtype (hvis denne er synlig i små strukturer), og noe større former som flagg eller virvlene i regnbuebildet. Den favner altså ikke stil i sin fulle betydning, men viktige deler av den. Antydningen fra Gatys, Ecker og Bethge om at nevral stiloverføring kan fortelle oss noe om hvordan vi som biologiske vesener oppfatter estetikk, er

noe spekulativ, selv om den kan se ut til å fange noe av de samme funksjonene som hjernen gjør ved objektsgjenkjenning. Hvordan fenomener som *mening* i innholdet kommer inn i denne modellen er ukjent for oss. Til slutt, bruksområdene til metoden for nevral stiloverføring har etter våre forsøk noen begrensninger, om det så skyldes vår utførelse eller metoden i seg selv. Hovedfunnet vårt er at bilder med mye representasjonelt innhold ikke egner seg like godt som stilbilde, og at rent estetisk har funksjonen dermed likheter med et filter. Vi foreslår dermed at metoden mest er konseptuelt interessant for kunstnere og vitenskapen.

Bibliografi

Carandini, M. 2012. “Area V1”. *Scholarpedia*. http://www.scholarpedia.org/article/Area_V1

Det Norske Akademis Ordbok (NAOB), s.v. “form”, lest 12. mai 2022, https://naob.no/ordbok/form.

Det Norske Akademis Ordbok (NAOB), s.v. “innhold”, lest 12. mai 2022, https://naob.no/ordbok/innhold.

Det Norske Akademis Ordbok (NAOB), s.v. “stil”, lest 12. mai 2022, https://naob.no/ordbok/stil.

Gatys Leon A., Alexander S. Ecker & Matthias Bethge. 2015. “A Neural Algorithm of Artistic Style”. *arXiv*. https://arxiv.org/abs/1508.06576

Kravitz, D. J. et al. 2013. “The ventral visual pathway: An expanded neural framework for the processing of object quality”. *Trends in cognitive sciences.* 17 (1): 26-49. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3532569/

Merriam-Webster.com Dictionary, s.v. “content,” lest 12. mai 2022, https://www.merriam-webster.com/dictionary/content.

Merriam-Webster.com Dictionary, s.v. “form,” lest 12. mai 2022, https://www.merriam-webster.com/dictionary/form

Merriam-Webster.com Dictionary, s.v. “style,” lest 12. mai 2022, https://www.merriam-webster.com/dictionary/style.

Neuralink. 2022. “Understanding the Brain”. Oppdatert 12. mai, 2022. https://neuralink.com/science/