

Emotionel billede analyse

Andreas Madsen s183901

Januar 2019

Abstract

Most people would agree that some pictures can evoke emotions. However classifying those emotions is often hard and very subjective, yet some features may give a picture an overall emotional value for a picture. Developing an effective emotional picture analyses model would serve many purposes in image search and art purposes. In this paper I test transfer learning as a method for developing a model. Even though my model achieved some accuracy (33%), it is easy to deceive. I think the reason for lack of accuracy is mostly do to a lack of data and bad labeling.

Indledning

Emotionel billede analyse går ud på at detektere følelser i billeder. Sådan et program ville give mulighed for at søge i billed databaser, give kunstnere en mulighed for at undersøge om deres kunst udsender den tænkte følelse og i sidste ende kan en AI tilgang muligvis skabe en bedre forståelse for hvad føleres er. Emotionel billede genkendelse lider under mangel på et stort datasæt, state of the art er derfor at hente billeder fra nettet der ligner træning sættet og træne med dem [3] vil jeg gøre brug af transferlearning i stedet for at træne et neuralt netværk fra bunden. Mit forsøg tager udgangspunkt i et andet forsøg lavet af Jana Machajdi og Allan Hanbury [1] som i 2010 byggede en classifier baseret på features fra kunstteori. Jeg har adgang til 2 ud af de 3 datasæt de brugte. Derfor har jeg i mit forsøg også de samme 8 følelses klasser som de har. De 8 klasser er følgende: Amusement, Anger, Awa, Contentment, Disgust, Excitement, Fear og Sad.

Hypotese

Gennem transferlearning kan et convolutional neural network(ResNet18) trænes til at genkende hvilken følelse/stemning et billed udstråler.

Etik

Affective computing har været etisk diskuteret i forhold til privatliv da følelser af mange betragtes som private[4]. At udvikle bedre metoder for computere til at forstå mennesker er derfor også at udvikle bedre metoder for firmaer at analysere deres brugere. De fleste problemer kan dog overses i dette projekt da mit projekt

handler om at forstå følelser generelt og ikke om at afkode enkelte personers følelsesliv. Dog ville teknologien give sider som Instagram og Diviant art, nye muligheder for at analysere en enkelt brugers konto for tegn på depression eller vrede.

Metode

Jeg har prøvet at gøre mit forsøgs opsæt så generelt som muligt så det tillader at teste flere modeller og datasæt. Dog ville det kræve et nyt dataloader modul hvis andet end pytorch skal bruges. Jeg har dog kun haft tid til at teste en model på et datasæt.

Min kode kan findes på følgerne git repository <https://lab.compute.dtu.dk/s183901/emotionel-biled-analyse.git>

Pilotforsøg

Jeg lavede et pilotforsøg med samme transferlearning model[2] men kun to output Happy og Sad. Det opnåede en accuracy på 92% på valideringssættet.

Data

Jeg har i forsøget brugt 3 datasæt. I mit pilotforsøg brugte jeg et datasæt bestående af billeder labeled Happy og Sad downloadet fra google image. Fra Machajdik og Hanbury har jeg et datasæt bestående af mellem 807 billeder i hver kategori. De havde også et dataset bestående af 280 stykker abstrakt kunst som var blevet fordelt ind i de 8 kategorier af 12 til 16 mennesker. Disse opdelinger var dog meget inkonsistente, jeg har derfor kun smidt de abstrakte billeder i en kategori hvis mere end fire mennesker havde valgt den, dette resulterede dog i at der var meget få billeder tilbage i nogle kategorier. Jeg har derfor i det endelige forsøg valgt at kombinere de to datasæt så jeg nu i alt har: 120 billeder labeled Amusement, 79 billeder labeled Anger, 116 billeder labeled Awe, 125 billeder labeled contentment, 84 billeder labeled Disgust, 123 labeled Exitement, 145 billeder labeled Fear, 189 billeder labeled Sad. Eksempler på billederne kan ses i appendiks ??.

Begge dataset kan hentes på <http://www.imageemotion.org/>

Classifyer

I mit forsøg har jeg brugt et færdiglavet convolutional neural net fundet på Pytorch hjemmeside [2]. Modellen er Resnet18 som er fortrænet på imagenet 1000 dataset. Jeg har 8 output neuroner en for hver følelse. Det sidste lag er trænet ud fra gradient nedstigning hvor alle andre lag end det sidste var fastlåste. Modellen træner 24 gange men returnerer modellen på det stadie hvor validering sættet opnåede den største accuracy på valideringssættet for at undgå over fitting.

Evalueringsmetode

Jeg har for at gøre op for manglerne data brugt k-cross validation. Hvert datasæt bliver delt op i fem dele. Jeg kører forsøget 5 gange hvor jeg træner en model på 80% datasættet og teste på den sidste femtedel. Træning sættet deler jeg op i

to dele hvor 20% procent (16% af hele datasættet) bliver brugt til at validere hvornår modellen over fitter og de resterne 80% (64% af hele datasættet) bliver brugt til at træne classificeren.

For hver af de 5 modeller udregner jeg en accuracy på hver af de 8 følelses kategorier og en middelværdi som er accuracy for hele modellen. Evalueringsmodulen returnerer også en confusion-matrix som kan give en ide om hvor fejlene opstår.

Når alle modellerne har kørt igennem vil jeg tage middelværdien for hver klasse og modellerne, samt en standardfejl som er givet:

$$\text{standard fejl} = \frac{1}{\sqrt{K}} \sqrt{\frac{1}{K} \sum_{i=1}^K (x_i - \bar{x})^2}$$

K er antallet af folder og i dette forsøg derfor =5 og x er accuracy for den enkelte model.

Jeg summere også alle konfusion matrixen blev lagt sammen for at få et større overblik over hvad der kunne være galt og ikke kun hvor den enkelte model laver fejl.

Resultater

kategori	middel accuracy	standard fejl
Amusement	48%	4%
Anger	10%	4%
Awe	20%	2%
Contentment	38%	4%
Disgust	26%	4%
Excitement	43%	6%
Fear	37%	4%
Sad	44%	3%
hele modellen	33%	3%

Table 1: Resultater

```

am ,an ,aw ,co ,di ,ex ,fe ,sa
[57.  6.  7.  8.  5. 22.  4. 11.]: amusement
[ 6.  8.  7.  7. 12.  5. 11. 23.]: anger
[15.  5. 24. 24.  3. 11. 14. 20.]: awe
[11.  4. 13. 48.  3. 11. 12. 23.]: contentment
[ 4.  7.  2.  5. 22.  5. 23. 16.]: disgust
[15.  2. 10. 11.  6. 54.  8. 19.]: excitement
[ 3.  6.  5.  2. 19. 12. 54. 45.]: fear
[13.  8. 14. 10.  6. 13. 41. 84.]: sad

```

Figure 1: fig:matix summeret over alle modeller

Diskussion

Resultater

Modellen er bedre end tilfældig hvilket ville være 12.5% accuracy og man må derfor sige at den kan genkende følelser på billeder selvom accuracy er lav. Hvis jeg sammenligner med Machajdi og Hanbury resultater kan vi også se at ren machinelearning ikke er bedre end kombineret machinelearning og kunstteori sammenlign figur 2 og mine resultater ??.

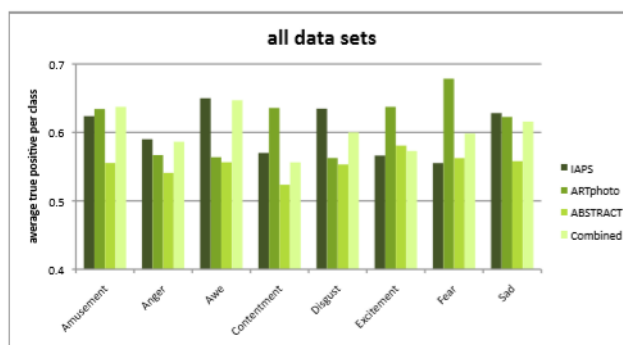


Figure 2: resultater opnået af Machajdi og Hanbury [1]

Data

Grundet datasættet størrelse vil jeg mene at hvis man skal forbedre accuracy skal man forbedre datasættet. Datasættets størrelse er klart et stort problem. Også det faktum at nogle dataset er større end andre er et problem, kigger man på konfusion matrixen kan man se at den har en tendens til at gætte på de store datasæt så som Fear og Sad. Dette problem kunne man programmere sig ud af ved at justere loss funktionen med datasættet størrelse. Ellers kunne man prøve at skabe lige store dataset. Et andet problem er labeling af billederne. Særligt datasættet fra Diviant art har nogen billeder som klart udstråler kunstnerens bias som for eksempel billede 3h som viser en subjektiv frygt for rutsjebaner selvom forlystelser hovedsageligt er blive labeled Exitement eller Amusement. Det samme kan siges om billede 3g som vil få vidt forskellige labels afhængig af personens holdning til koncerter. Data sættet med abstrakt kunst er evalueret af flere mennesker og er kun indgået i forsøget med label hvis der har været en enighed blandt deltagerne som i 3c. Mange billeder som billede som 3i blev sorteret fra i den sammenhæng.

I fremtidige forsøg vil jeg foreslå ikke at bruge ord som labels der kan misforstås af mange men i stedet prøve at følge tilgangen i IAPS datasættet [5] hvor man har prøvet at klassificere nogen grund føleres som vil være til stede i alle billeder og have en ikke binær værdi. IAPS bruger 3 kategorier Happy vs. Unhappy, Excited vs. Calm og Controlled vs. In control. Disse tre akser svare nogenlunde til 6 af kategorierne i mit forsøg, man kunne også lave en 4 akse kaldet like vs. dislike selvom den burde illustrer et meget subjektiv forhold til objektet på billedet.

For at uddybe det lingvistiske problem kan man sige at Happy, Unhappy svare til Amusement og Sad. Problemet er at Amusement kan ses som en kombination Happy og Exited for eksempel er billed 3f labeled Exitement men et billed af den samme type forlystelse finde i træningssættet som Amusement. Dette giver en stor forvirring for modellen, som har svært ved at skille de to ad se konfusion matrix 1. Et andet problem er at Amusement også kan forstås som om billedet er sjovt 3a at bygge en model der kan detekter humor må være et andet projekt.

Datasættet labeled anger var specielt problematisk da det ud over at være det mindste datasæt hovedsageligt havde billeder af vrede mennesker, vrede dyr eller som i billed 3b en vred natur. Derudover er vrede en føles der er svær at kategorisere i et statisk billede da vrede ofte er frembragt af en handling [6] Derudover kan anger klassificeres som en kombination af In-control, Unhappy og dislike af objektet anger er rettet imod. På samme måde kan Fear beskrives som en kombination af Controlet, Unhappy og Dislike. Derfor er det også let at bytte rundt på anger og fear selvom de er modsætninger på Contol skalaen. At modellen har problemer med at skelne mellem Anger, Fear, Disgust og Sad ses tydeligt i confusion matrixen Figurematix

Problemet med Awe og Disgust er at de omhandler en meget subjektiv holdning til objektet på billedet. Dog vil de fleste mennesker mene at nuttede dyr, natur, og rene mennesker 3d er i Awe og blod 3e, samt beskidte mennesker og dyr er i Disgust. Awe og Disgust var noget af det modellen havde sværest ved at detektere dog viser konfusion matrixen at de sjældent bliver byttet helt om 1.

Andre evalueringsmetoder

et forslag til et andet eksperiment må derfor være at sammenligne resultaterne i hver kategori med dens modsætning og observer hvad der kommer tættest på labelen. Dette gjorde jeg i mit pilotforsøg det er også måden andre har evalueret machinelearnings modeller [3]

I stedet for k-fold kunne man også evaluere usikkerheden på en enkelt model gennem konfidentiel intervallet. Jeg regnede mig frem til at for at opnå en usikkerhed på max 10% med et konfidentiel interval på 90% skulle jeg bruge 30 billeder. Da mit mindste dataset har 77 billeder virkede dette ikke realistisk. Jeg har lavet en funktion til at udregne usikkerheden på den enkelte model men bruger den ikke, da jeg bruger standard fejlen i stedet.

Konklusion

Jeg har opnået emotionel genkendelsesglæde med 33% accuracy gennem transferlearning.

References

- [1] Jana Machajdik: Allan Hanbury *Affective image classification using features inspired by psychology and art* <http://www.imageemotion.org/>
- [2] Sasank Chilamkurthy https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html?fbclid=IwAR2k1y-mvVyU4F6uiwAv1TPXZ3Qvnp0MabCcaA-RrzZmKtFn5pNnuw6D3gg
- [3] Na Li: Yong Xia Affective image classification via semi-supervised learning from web images
- [4] Carson Reynolds: Rosalind Picard Affective Sensors, Privacy, and Ethical Contracts
- [5] Lang, P.J: Bradley, M.M: Cuthbert, B.N: International Affective Picture System (IAPS): Technical Manual and Affective Ratings
- [6] JOSEPH A. MIKELS: BARBARA L. FREDRICKSON: GREGORY R. LARKIN: CASEY M. LINDBERG: SAM J. MAGLIO: PATRICIA A. REUTER-LORENZ. Emotional category data on images from the International Affective Picture System

Appendiks



Figure 3: Eksempler på prediction