



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI PENELITIAN

3.1. Pengumpulan Data

3.1.1. Sumber Data

Data yang digunakan, ialah data yang juga digunakan pada penelitian *Pooling in image representation: The visual codeword point of view* oleh dr. Sandra Avilla. Pemilihan dalam penggunaan data ini didasarkan pada objek penelitian yang sesuai. Serta sudah memiliki kelas yang sudah didefinisikan, sehingga menghemat sumber daya jika dibandingkan dengan mencari data secara mandiri, kemudian menentukan kelas pornografi atau non-pornografi.

3.1.2. Proses Permohonan Data

Proses permintaan data relatif singkat, yakni dengan melakukan penandatanganan surat perjanjian yang telah disediakan sebelumnya, kemudian mengirim *email* beserta surat perjanjian tersebut.

NPDI Pornography Database License Agreement

This agreement is made by and between:

SINGGIH SURYA DHARMA

(hereinafter called END-USER), having its principal place of business at:

UNIVERSITAS MULTIMEDIA NUSANTARA - INDONESIA

AND

NPDI group, Federal University of Minas Gerais (UFMG), Computer Science Department, Belo Horizonte, Minas Gerais, Brazil, contact Sandra Avila Tel. +55 19 35 21 03 34.

Disclaimer

The NPDI Pornography Database IS PROVIDED "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. The videos, segments and images provided were produced by third-parties, who may have retained copyrights. They are provided strictly for non-profit research purposes, and limited, controlled distributed, intended to fall under the fair-use limitation. We take no guarantees or responsibilities, whatsoever, arising out of any copyright issue. Use at your own risk.

Release of the Database

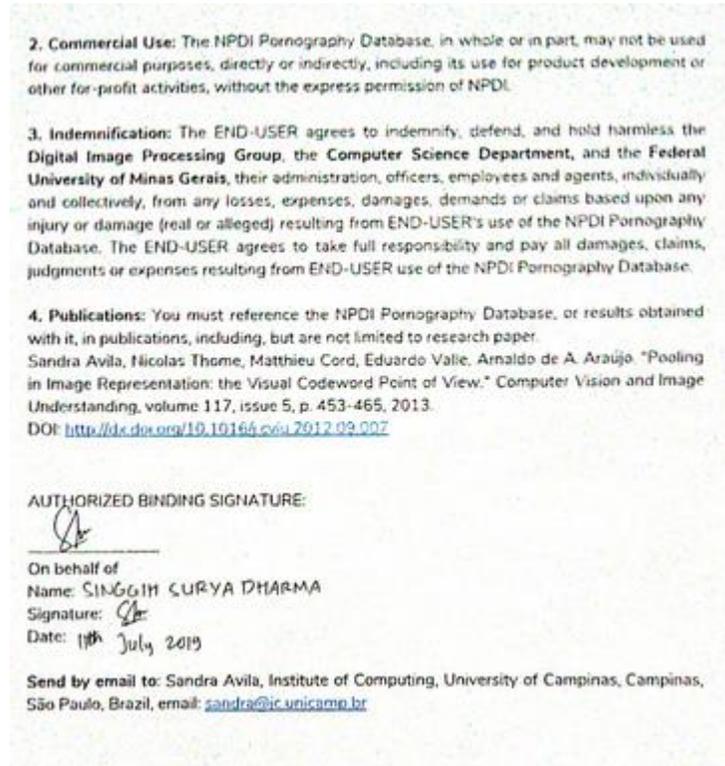
The NPDI Pornography Database is available to researchers under request, which will be evaluated case-by-case. All requests for the NPDI Pornography Database must be submitted to Prof. Sandra Avila. To receive a copy of the database, END-USER named must be a legal institution, or a department or section of a named legal institution, not an individual nor a project – the person signing this Agreement has to be duly authorized by the institution for such signatures (e.g., Department or Administrative Head or similar) and shall be liable for such authorization.

Terms and Conditions

By signing this document, END-USER agrees to receive a copy of the NPDI Pornography Database, under the following restrictions and obligations:

1. Redistribution: the NPDI Pornography Database cannot be, in whole or in part, further distributed, published, copied, disseminated or broadcast, in any form, whether for profit or not. That includes further distribution to a different facility or unit within the END-USER's university, organization, or company. An exemption is provided for a small number of individual static frames of the videos (less than 50 on any individual publication), which can be used for illustrations of scholarly posters, presentations and publications.

Gambar 3.1. Surat Kontrak Permohonan Data



Gambar 3.2. Surat Kontrak Permohonan Data (2)

Komposisi data yang didapatkan ialah sebagai berikut:

Table 3.1. Komposisi Data Awal Keseluruhan

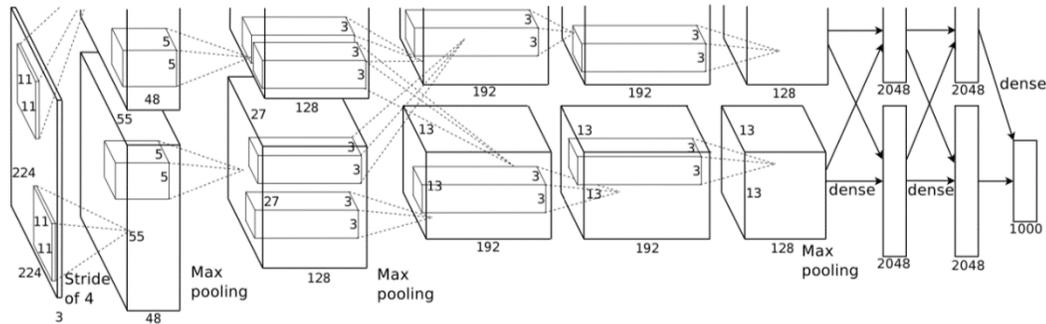
Class	Videos	Hours	Shot per Video
Porn	400	57	15.6
Non-porn(easy)	200	11.5	33.8
Non-porn(difficult)	200	8.5	17.5
All videos	800	77	20.6

Table 3.2. Komposisi Data Awal Keseluruhan(2)

Ethnicity	% of Videos
Asians	14%
Blacks	16%
Whites	46%
Multi-ethnic	24%

3.2. Metode Penelitian

Penelitian ini akan melihat cara kerja dari layer-layer yang ada pada metode *deep learning* pada suatu model yang sering kali dianggap *blackbox*. Dalam penelitian ini, model yang akan digunakan ialah AlexNet dengan struktur model sebagai berikut.



Gambar 3.3. Model AlexNet

Sumber: (Krizhevsky et al., 2012)

Tahapan penelitian yang dilakukan setelah didapatkan data, dapat dilihat pada gambar 3.4.



Gambar 3.4. Tahapan Penelitian

3.2.1. Merubah Ukuran Semua Gambar (*Resize*)

Data yang telah disediakan ada berupa video, *frames*, maupun *segments*. Dari ketiga tersebut, penelitian ini menggunakan data yang berjenis *frames/gambar*. Karena ukuran gambar yang disediakan semuanya berbeda, gambar harus diubah dahulu menjadi 224 x 224 x 3. Pengubahan ukuran ini dilakukan dengan menggunakan PIL.

3.2.2. Melakukan Fold Data dengan 10-Fold Cross Validation

Penggunaan *k-fold cross validation* dengan K bernilai 10 digunakan penelitian ini untuk menguji kekuatan dari model atas data tersebut. Penggunaan *cross validation* ini dilakukan dengan mengambil 10% dari data secara acak dan menjadikannya data *test* dan sisanya sebagai data *train*

menggunakan fungsi *random* dari Python. Pengujian akurasi dilakukan untuk sepuluh *fold* yang sudah dibuat.

3.2.3. Mengubah Gambar menjadi Array

Pengubahan ini gambar menjadi *array* dilakukan dengan menggunakan *library* OpenCV. Tiap gambar diubah menjadi *array*, kemudian *array* gambar ditambahkan menjadi sebuah *array* lagi.

3.2.4. Melakukan Labelisasi

Penentuan *fold* sudah dilakukan, maka perlu juga disiapkan label dari tiap data pada masing-masing *fold* baik data *train* maupun data *test*. Untuk kelas porno, diberikan label [0,1] dan label [0,0] untuk non-porno. Proses labelisasi dilakukan dengan python menggunakan *library* NumPy dengan bantuan LibreOffice, Penggunaan LibreOffice digunakan untuk melakukan *if conditional* serta memberikan label sesuai dengan kelas masing-masing, rangkaian yang dihasilkan pada LibreOffice akan disusun menjadi *array* dengan NumPy.

3.2.5. Melakukan Training pada Tiap Fold dan Melihat Akurasinya

Training yang dilakukan untuk setiap *fold*, dilakukan dengan model AlexNet yang sudah dijelaskan strukturnya sebelumnya. Proses *training* dilakukan untuk melihat akurasi antar *fold*. Jika ada perbedaan akurasi yang signifikan, maka harus dilihat hasil gambar pada layer di setiap *fold* tersebut.

3.2.6. Menentukan 1 *Fold* yang akan digunakan

Ketika akurasi antar *fold* tidak jauh berbeda, maka digunakan salah satu *fold* saja untuk dilihat hasil di tiap layernya. Karena tiap *fold* memiliki kesamaan. Kesamaan dari akurasi tiap *fold* akan diukur dengan signifikansi menggunakan metode ANOVA. Pemilihan secara acak dilakukan dengan bantuan fungsi *random* dari python untuk menentukan *fold* mana yang digunakan.

3.2.7. Membandingkan Beberapa Hasil Gambar dari Tiap Layer pada *Fold*

Untuk mengetahui, sudah sesuai dengan model yang ditentukan, maka diujikan beberapa gambar untuk melihat kesesuaian tiap layer. Perbandingan dilakukan pada gambar dari kedua kelas, baik gambar porno dan non-porno.

3.2.8. Melakukan *Masking* pada Layer Pertama dengan *Source* Gambar

Proses ini sama seperti proses *masking* antar hasil dari tiap layer, namun bedanya proses *masking* ini dilakukan pada gambar awal (*source*) dengan gambar hasil dari konvolusi layer 1. Gambar dari hasil konvolusi layer 1 memiliki sumbu x dan sumbu y, maka untuk melakukan *masking*, gambar tersebut harus dipotong beberapa bagiannya, sehingga mampu dilakukan *masking* dengan hasil yang lebih jelas.

3.2.9. *Masking* antara Hasil Gambar dari Tiap Layer

Masking tiap layer dilakukan dengan OpenCV, proses *masking* dilakukan dari gambar sebelum melewati layer dengan gambar sesudah melewati layer. Dengan kata lain, hasil dari konvolusi layer 1 akan dibandingkan dengan hasil konvolusi layer 2, dikarenakan hasil dari konvolusi layer 1 adalah gambar sebelum melewati konvolusi layer 2. Karena dalam 1 layer terdapat beberapa *channel*, maka dilakukan rata-rata dari *channel* dari 1 layer tersebut.

3.3. Environment Penelitian

Penelitian yang dilakukan, berlangsung pada komputer server milik program studi Sistem Informasi Universitas Multimedia Nusantara, berikut spesifikasi dari komputer server yang digunakan terkait dengan penelitian:

Table 3.3. Spesifikasi Komputer untuk Penelitian

Jenis	Keterangan
Sistem Operasi	Ubuntu 18.04.3 LTS (Bionic Beaver)
Kartu Grafik	GeForce GTX 1080
Driver Kartu Grafik	Driver Version: 430.26 CUDA Version: 10.2
Processor	Intel Core i7- 8700 @ 3.20GHz 6 Cores 12 Threads
Memory	32 GB , 8GB (swap memory)
Python	Python 3.7.3

Proses penelitian model dilakukan dengan menggunakan *command line* linux, dengan *text-editor* sebagai aplikasi untuk menulis barisan kode.