

CONFIDENTIAL UP TO AND INCLUDING 31/12/2032 - DO NOT COPY, DISTRIBUTE OR MAKE PUBLIC IN ANY WAY

Using GANs to optimize Pose Estimation on Art Collections

Tristan Verheecke

Student number: 20043518

Supervisors: Prof. dr. ir. Dieter De Witte, Prof. dr. Steven Verstockt

Counsellor: Kenzo Milleville

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Information Engineering Technology

Academic year 2022-2023

Preface

I've been interested in Art my entire life. In fact, I've a degree in the Fine Arts from LUCA School of Arts. There, I was known for my technical ability and one of my professors at the time asked me why I didn't do anything with that in my artworks. That remark has since stuck with me and was part of my motivation to apply for readmission for my Master of Science. With all the advancements in AI, I started thinking more and more about doing work with that. Like Matisse and Turner, I'm not satisfied with the tools available, but want to create my own.

It was therefore to my delight that I was able to work on this thesis which has provided me the opportunity to acquire more insight in the subject. I would like to thank my supervisors Dieter De Witte and Steven Verstockt for this wonderful opportunity, and my counsellor Kenzo Milleville for his great guidance. As well as all the other people at IDLab for their feedback. I also want to thank Karine Lacaracina, Lies Van De Cappelle and the other people at RMFAB for providing help with the artistic sensibilities of the thesis.

Enjoy the read,

Tristan Verheecke
Ghent, June 2023

Conference Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Abstract—This document is a model and instructions for L^AT_EX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

This document is a model and instructions for L^AT_EX. Please observe the conference page limits.

II. EASE OF USE

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections III-A–III-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—L^AT_EX will do that for you.

Identify applicable funding agency here. If none, delete this.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”).

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not

“Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

D. *L^AT_EX-Specific Advice*

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in L^AT_EX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

BIBT_EX does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use BIBT_EX to produce a bibliography you must send the .bib files.

L^AT_EX can’t read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

L^AT_EX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.

- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^aSample of a Table footnote.



Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
- [2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, “Title of paper if known,” unpublished.
- [5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.

Contents

List of Figures	viii
List of Tables	ix
List of Acronyms	x
List of Code Fragments	xiii
1 Introduction	1
1.1 Problem definition	1
2 Literature study	2
2.1 Human Pose estimation	2
2.1.1 Representation	2
2.1.2 Datasets	3
2.1.3 Discriminative Methods and Generative Methods	4
2.1.4 Single-Person Methods	4
2.1.5 Multi-Person Methods	6
2.2 Image Style Transfer	7
2.2.1 Datasets	7
2.2.2 Optimization-based Networks	11
2.2.3 Feed-forward Generation Networks	11
2.2.4 Generative Adversarial Networks	14
3 Titel tweede hoofdstuk	21
3.1 Sectie titel	21
4 Titel derde hoofdstuk	23
4.1 Sectie titel	23
4.2 Sectie titel2	24
Subtitel	24
4.2.1 Functie is_isbn	25
Conclusie	26

Ethische en maatschappelijke reflectie	26
Referenties	27
Bijlagen	33
Bijlage 1	34
Bijlage 2	35

List of Figures

2.1	The various challenges HPE solutions face. Images from Max Planck Institute for Informatics (MPII) dataset. [1][2]	8
2.2	Models for pose representation [3]	8
2.3	The different methods of single-person human pose estimation.[3]	9
2.4	Convolution layers in blue and fully connected layers in green. The initial stage is applied to the whole images, while in stage s it will work on a sub-image based on the result of the previous stage.[4]	10
2.5	Architecture and receptive fields of Convolutional Pose Machines (CPMs). (a) and (b) represent the pose machine architecture.[5] (c) and (d) show the corresponding convolutional networks used by CPMs.[6]	11
2.6	The structure of a "stacked hourglass" network and a single "hourglass" module.[7]	12
2.7	The architecture of the High-Resolution network and how it applies multi-scale fusion.[8]	13
2.8	Cascaded Pyramid Network. "L2 loss*" means L2 loss with online hard keypoints mining.[9]	13
2.9	Style transfer algorithm. (Gatys et al. [10]).	15
2.10	A texture network by Ulyanov et al. [11]. The generator network (left) is the only one that changes. The descriptor network (right) is used to calculate the loss.	16
2.11	An image transformation network by Johnson et al. [12]. The image transform network (left) is the only one that changes. A loss network (right) is used to define perceptual loss functions.	16
2.12	A comparison between (c) Batch Normalization (BN) and (d) Instance Normalization (IN).[13]	16
2.13	A comparison between different style transfers where the style was not seen during training.	17
2.14	The stylebank network by Chen et al. [9].	17
2.15	The domain transfer network by Taigman et al. [14].	18
2.16	The cycle-consistent network.	19
2.17	(a) The shared latent space assumption. (b) The unsupervised image-to-image translation network by Liu et al. [15].	20

List of Tables

List of Acronyms

A

AdalIN	Adaptive Instance Normalization , 14
AIC-HKD	AI Challenger Human Keypoint Detection , 4
ASMs	Active Shape Models , 3

B

BN	Batch Normalization viii, 11, 16
----	----------------------------------

C

cGAN	conditional Generative Adversarial Network , 5, 14
CIN	Conditional Instance Normalization , 14
CNN	Convolutional Neural Network , 3, 5–7
COCO	Common Object in Context , 3, 7
CPMs	Convolutional Pose Machines viii, 5, 7, 11
CPN	Cascaded Pyramid Network , 6

F

FLIC	Frames Labeled In Cinema , 3
------	------------------------------

G

GAN	Generative Adversarial Network , 5, 14, 15
-----	--------------------------------------------

H

HPE Human Pose Estimation , 2–5, 7

I

ILP Integer Linear Programming , 6
IN Instance Normalization viii, 11, 16

L

LSP Leeds Sports Pose , 3

M

MPII Max Planck Institute for Informatics viii, 3, 4, 8
MSE Mean Square Error , 14

N

NMS Non-Maximum-Suppression , 6
NST Neural Style Transfer , 14

P

PAF Part Affinity Field , 7
PAF Part Association Fields , 7

PIF

Part Intensity Fields , 7

R

ResNet

Residual Network , 5, 6

RMFAB

Royal Museums of Fine Arts of Belgium , 1

RPME

Regional Multi-person Pose Estimation , 6

S

SAHR

Scale-adaptive Heatmap Regression , 7

SMPL

Skinned Multi-Person Linear , 3

V

VAE

Variational Autoencoder , 15

W

WAHR

Weight-adaptive Heatmap Regression , 7

List of Code Fragments

4.1 Functie is_isbn	25
------------------------------------	----

1

Introduction

1.1 Problem definition

To make art collections more accessible, museums put a huge effort in digitalizing their catalogue. However, they don't contain much metadata about the content and it is time-consuming to enhance them manually. To make this process easier, they want to utilize computer vision. Art collections (paintings, statues, drawings, etc.) turn out to be less interpretable by the algorithms that were developed for photography over the last few decades. These scan the images in search of recognizable objects and add their labels to the metadata. Even the latest state-of-the-art technology, struggles to recognize objects when pointed at a painting in a museum. A solution may be to start over and have paintings annotated by humans.

This has been done in 2 recent projects: Saint-George-On-A-Bike [16] and INSIGHT [17]. However, paintings are very complex and manual annotation doesn't scale and is very expensive. For example, 10,000 paintings were annotated by Royal Museums of Fine Arts of Belgium (RMFAB) with no clear return on investment. They spent a year on this and this is not something they want to repeat. How can we automate this process and ensure that state-of-the-art computer vision models give good results on paintings and artworks?

Specifically for this thesis, pose estimation will be investigated.

1.2 Proposed solution

We will first examine the effectiveness of existing models on a collection of paintings from 2 different movements. For this we will need to have a pose estimator, a style transformer and a collection of test data.

A first method: We will first convert the test data with the style transformer to a painting and then we will apply pose estimation. The test data will have coordinates of the joints, which we will compare with the results of the pose estimation. However, the joints are of the original image. How do we convert those coordinates to map to the styled image? Problem: This method does not use any real paintings and will be susceptible to the accuracy of the style transformer.

A second method: We can apply pose estimation to real paintings and then convert them to a realistic image with style transfer. We can then use pose estimation to the realistic images and compare them with the style transformed results. This will also require a way to map the results of the real painting to that of the style transformed. Problem: While we're using real paintings now, the results will still depend on the accuracy of style transformer.

A third method: We can annotate the paintings ourselves and use pose estimation to assess the pose estimation algorithms. Problem: We must annotate the paintings ourselves.

There are several things that can be improved: The dataset, the algorithm, the input

2

Literature study

In order to correctly implement a solution, we need to understand the fundamentals. These consist of 2 research fields: Human Pose Estimation (HPE) and Neural Style Transfer (NST). The former will be used to detect poses in the art collections, but not before the latter has tried to make an improvement. Following will be an overview of the available research in these domains. Discussing what the goals of them are, how they achieve it and their limitations.

2.1 Human Pose estimation

HPE aims to detect human features from input data such as images and videos. It's an elementary part of computer vision with many applications among which are human action recognition (sign language), human tracking (surveillance), and human-computer interaction (video games). This is an extensively researched area with a diverse range of different techniques. This chapter will try to give an overview of all the many challenges and proposed solutions. The focus will be on deep learning models, which have surpassed classical solutions significantly. Specifically, around 2D monocular HPE eg., [18][3][19][20].

The human body has a high degree-of-freedom due to all the limbs, self-similar parts and body types, which may cause self-occlusion or rare/complex poses. The variations in configuration are made even larger due to clothing, lighting, foreground occlusion, as well as viewing angles and truncation, among others, as shown in fig. 2.1. This makes HPE one of the most difficult tasks in computer vision [21][2].

2.1.1 Representation

An important factor in HPE is how the pose will be represented. Depending on the needs of the problem you can have a skeleton-base, contour-base, or volume-base solution [2]2.2.

Skeleton-based model

The skeleton is build of a tree-structured set of keypoints that represent the joints of the human body. These can be explicitly described by their coordinates in 2D or 3D space [4]. More suitable for a Convolutional Neural Network (CNN) however is a heatmap which constructs a 2D Gaussian kernel around a keypoint [19][22]. They are easily implemented and became the dominant representation. While the skeleton-based model is a compact and flexible representation it suffers in this aspect by not being able to hold texture or shape information [3].

Contour representation

To capture the shape of the body parts, contour representation uses rectangles to estimate the body contours. These methods include cardboard models [23] and Active Shape Models (ASMs) [24] and were mainly in use in earlier HPE methods [2].

Volume representation

Volumetric geometric shapes can also be used as a method of representation. Earlier methods used simple shapes like cylinders, conics, and other shapes [25]. Volume representation is a 3D mesh that represents the human body. The most used model is Skinned Multi-Person Linear (SMPL), which includes natural pose-dependent deformations imitating soft-tissue dynamics [26].

For the purpose of our research, a simple model is the only thing we need. We only need to be aware of the most essential joints to label a pose. This makes the skeleton-based model the ideal representation to work with and will be the focus of further study.

2.1.2 Datasets

There are several publicly available datasets. There are some that are outdated and we will leave those out, focusing only on datasets used for deep learning.

Leeds Sports Pose (LSP) Dataset [27] contains 2,000 images found on Flickr using 8 different tags looking for sport activities (athletics, badminton, baseball, gymnastics, parkour, soccer, tennis, and volleyball). Each person has 14 keypoints. An extended version was later introduced [28], now consisting of 10,000 images. For this set they only focused on the more challenging tags (parkour, gymnastics, and athletics).

MPII Human Pose Dataset [1] contains 24,290 images with 40,522 labeled people. They were extracted from YouTube videos found by querying for physical activities. Each person has 16 keypoints and also includes occlusion labels.

Common Object in Context (COCO) Dataset [29] is a large-scale dataset for a wide range of computer vision algorithms. For HPE, the set contains more than 200,000 images in which 250,000 persons are annotated. Each person has 17 keypoints, a bounding-box and visibility labels. This dataset has become the most popular for benchmarking.

Frames Labeled In Cinema (FLIC) Dataset [30] contains 5,003 images extracted from Hollywood movies. They ran a person detector which collected 20,000 images from 30 movies. Occluded and difficult poses were then removed leaving only 5,000 images to be annotated. Only the upper body received 10 keypoints.

AI Challenger Human Keypoint Detection (AIC-HKD) Dataset [30] contains 300,000 images found using Internet search engines. In these, over 700,000 humans are annotated. Each person has 14 keypoints, a bounding-box, as well as visibility and left/right labels.

CrowdPose Dataset [31] puts an emphasis on crowded images. 30,000 images from MPII, glsCOCO and glsAIC-HKD were measured with a Crowd Index, which evaluates the crowdedness. Finally, 20,000 images are selected and 80,000 persons annotated. Each person has 14 keypoints and a full-body bounding box.

2.1.3 Discriminative Methods and Generative Methods

Before deep learning became prominent in HPE there were already a number of different methods in use. Some of these methods are compatible with the deep learning methods and were thus adopted. An early distinction is between generative and discriminative methods.

Generative Model

A generative method will work with prior beliefs about the pose. More information about this can be found in the section about representation 2.1.1. It will project the pose on the image and verify it with the image data. If they don't comply, the pose is adjusted using the descent direction found by minimizing an error function [32].

Discriminative Model

Discriminative methods on the other hand, try to map the pose on the image data with learned models. There are several methods in this category, among which are the deep learning-based methods. The deep-learning methods are further categorized by the following sections.

2.1.4 Single-Person Methods

Single-person pose estimation will try to evaluate only one pose from an image. There are 2 major methods that are in use: regression methods and detection-based methods.

Regression-based Methods

The regression-based methods learn a network that maps all the body keypoints to the image-data directly as shown in 2.3a.

The first successful deep learning model came from Toshev and Svededy [4] and is considered the switch in paradigm from classic approaches to deep learning HPE. Toshev et al. uses a 7-layered model with 5 convolution layers and 2 fully-connected layers for the pose regressor, based on AlexNet for its simple but effective architecture [33]. They then cascade the resulting found keypoints of this model to itself where it refines it using the area around the keypoints. While the network is the same, the different stages will have different learned parameters. With every stage the found keypoints become more accurate.

Carreira et al. [34] introduce an Iterative Error Feedback which is a self-correcting model using top-down feedback. Using the image-data and a starting pose modeled as a heatmap, the model, based on GoogLeNet [35], will predict an error for each keypoint. The pose is then corrected based on the error and fed back into the model as a heatmap with the image. With each iteration it converges towards the solution instead of making the prediction in one go. Regression-based methods map the keypoints directly on the image, making it a non-linear problem. This will cause less robust generalization [19].

Heatmap/Detection-based Methods

The detection-based methods will first estimate the individual body parts using heatmaps, which leads to an easier optimization and a more robust generalization [20]. Most of the latest HPE methods use heatmaps because of this. After the

joints are found they are then assembled to fit a human skeleton. This process is shown in 2.3b.

Tompson et al. [36] proposed a hybrid architecture where the detection of body parts is handled by a CNN and a Spatial-Model to bring those together. The first step produces many false-positives and these are removed in the second step by restricting joint inter-connectivity to enforce correct anatomy. They build on this in [37], where they used a cascade to refine predictions.

A fundamental work written by Wei et al. [6] combines convolution networks with Pose Machines [5]. Pose Machines is an iterative architecture which consists of 2 models: the first is used for stage 1 where it extracts potential heatmaps for the joints. The second model is used for subsequent stages where the result of the previous stage is fed in together with the results of its own convolution network on the input image. This gradually refines the predictions for the joints and their positioning. 2.5 shows this process.

Another influential work was being written at the same time by Newell et al. [7]. Similar to CPMs, this is also an iterative architecture. They suggest what they call a "stacked hourglass" network, where "hourglass" modules are repeated 2.6a. In an "hourglass" module, first, the features are downsampled and afterwards upsampled again 2.6b. This network captures different spatial relationships between joints at different resolutions. Several other works [38][?][39] have since improved on the network design.

Both these use intermediate supervision to tackle the problem of vanishing gradients. This still doesn't build a deep sub-network for feature extraction which limits the estimations. This has become less of a problem with the emergence of Residual Network (ResNet)[40] which allows better back-propagation at deeper levels through shortcuts.

A more recent work by Sun et al. [8] maintains the high-resolution representations instead of working the high-resolution from the low-to-high sub-network. After a first high-resolution sub-network, it gradually adds high-to-low sub-networks in parallel to predict multi-resolution features. Before each branch, they apply multi-scale fusion, which joins the predicted features from each scale on each scale. Both are shown in 2.7. This network has proven very effective and inspired several variations [41][42][43].

With the emergence of neural networks also came Generative Adversarial Networks (GANs) [44], which proved useful for HPE. They are employed to improve constraints of joint inter-connectivity and infer occluded body parts.

Chen et al. [9] propose a structure-aware convolution network using a stacked hourglass as generator which generates heatmaps for each joint. They use 2 discriminators, one to discriminate between low- and high-confidence predictions, another for real and fake poses. The network is designed as a conditional Generative Adversarial Network (cGAN) [45], which allows it to generate pose heatmaps as well as occlusion heatmaps.

A more classic GAN is used by Chou et al. [46], where they use a stacked hourglass network for both the generator as the discriminator. The generator predicts the heatmaps for each joint and the discriminator distinguished between the real and fake ones.

2.1.5 Multi-Person Methods

With multi-person methods comes an extra layer of difficulty: they need to be able to detect each person separately. To solve this problem multi-person methods propose several solutions. The 2 most popular are top-down and bottom-up methods.

Top-Down Methods

This method will first try to detect all persons in the image with a human detector. Each person is cropped by the bounding box and a single-person estimator predicts a pose for each person.

Occlusion and truncation are a regular occurrence in multi-person scenes and inevitable problem. One of the early multi-person models, by Iqbal et al. [47], works towards creating a robust model against occlusion. It uses Faster RCNN [48] to detect the human boundaries. After which, it applies integer linear programming for each person's fully connected graph. This technique is similar to [49], but instead of working on all globally found joints it only considers local joints. It can also handle any kind of occlusion or truncation.

The use of a human detector comes with its own sort of problems. Fang et al. [50], with Regional Multi-person Pose Estimation (RPME), try to remedy these with 2 components: They try to tackle inaccurate bounding boxes with Symmetric Spatial Transformer Network, redundant detections with Parametric Pose Non-Maximum-Supresion. They also propose a 3rd component, Pose-Guided Proposals Generator, which can augment training samples.

Papandreou et al. [51] use a 2 stage pipeline. In the first stage, they employ the Faster RCNN detector [48]. In the second stage, they estimate the pose in each found bounding box using their own network. It predicts heatmaps using a fully convolutional ResNet and use their own novel aggregation procedure. Afterwards, they do post-processing using keypoint-based Non-Maximum-Suppression (NMS) a method of their own making.

A continuous effort is taken by Chen et al. [9] to deal with occlusion and truncation. They suggest a 2 stage architecture, a Cascaded Pyramid Network (CPN) as seen in 2.8, where first the "simple" keypoints are captured with GlobalNet, a feature pyramid network based on [52], and the "hard" keypoints are handled by their RefineNet, based on the upsampling and concatenating of HyperNet [53] and using an adapted stacked hourglass. They achieved great results and several others improved on their work [54][55].

In more recent research, a new method was become more powerful than CNNs. The Transformer [?], based on attention mechanisms which are used to optimize recurrent networks [56], eliminates the use of recurrent layers, keeping only the attention mechanisms. Yang et al. [57] use this architecture because allows for better understanding of the spatial dependencies and learns at a higher rate.

Bottom-Up Methods

A different approach is taken with bottom-up methods. They first locate all joints in the image and then assemble them in potential humans.

DeepCut by Pishchulin et al. [49], one of the first multi-person models using CNNs. Using Fast R-CNN [48], it detects the body parts and labels each. With the joints found, it then uses Integer Linear Programming (ILP) to assemble them. This method is very computationally expensive; NP-hard. Insafutdinov et al. [58] therefor introduce a stronger part detector and better optimization strategy with DeeperCut.

CPMs make a return with OpenPose by Cao et al. [59], they're used to predict the joints with heatmaps and Part Affinity Fields (PAFs). A part affinity field also encodes the position and orientation of the limb which makes the assembly of joints into different poses possible. They can achieve real-time results with this method, and several others have improved on their design [60][61][55]. The high performance is only applicable to high-resolution images. Low-resolution images or images with occlusions perform poorly.



Figure 2.1: The various challenges HPE solutions face. Images from MPII dataset. [1][2]

Kreiss et al. [62] continue on the idea of fields and introduce the Part Intensity Fields (PIF) and Part Association Fields (PAF). First, they predict the location of the different joints with PIF. Afterwards, they use PAF to find the inter-joint relationships. They are able to outperform any previous OpenPose-based proposals on low-resolution and occlusions.

Newell et al. [63] introduce a new method called associative embedding for supervising CNNs both detection and grouping. This is a single-stage architecture as opposed to the two-staged architectures previously discussed. They make use of the stacked hourglass network from [7] with some small modifications.

Continuing on the idea of associative embedding, Cheng et al. [41] use HRNet [8] as backbone for their HigherHRNet. Their method focuses on the scale-variance problem; a problem which hasn't been studied much, so it can localize keypoints for small persons better. Lou et al. [64] introduce Scale-adaptive Heatmap Regression (SAHR) and Weight-adaptive Heatmap Regression (WAHR) to the scale-variance problem. SAHR adaptively adjusts the standard deviation of each heatmap corresponding with the scale of the person. WAHR rebalances the foreground and background samples, so SAHR can work to its fullest extent.

Summary

An important challenge for HPE is making predictions in scenes with hight occlusions. Top-down models achieve state-of-the art performance in almost all benchmark datasets [2]. Top-down models has difficulty with overlapping bodies and human detectors might fail finding humans there. To the same extent, bottom-up models will have greater inaccuracy with grouping in occluded scenes. Computationally, the top-down model's speed is limited by the number of people found. The higher efficiency of bottom-up models, make them more suitable for real-time applications.

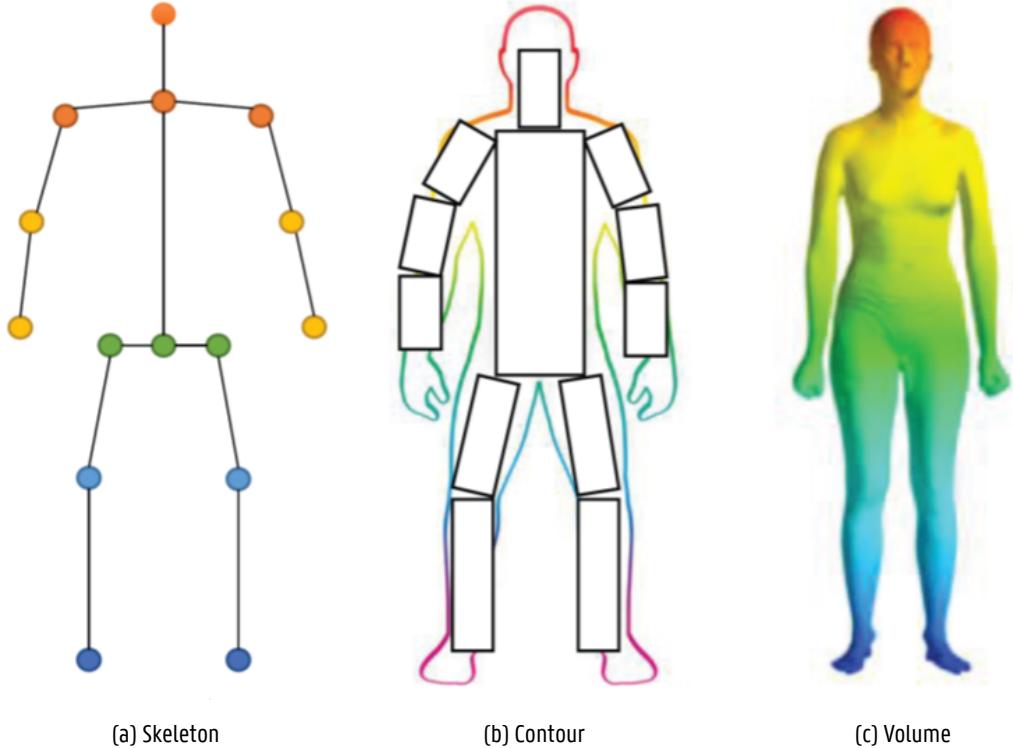


Figure 2.2: Models for pose representation [3]

2.2 Image Style Transfer

Image Style Transfer is the technique of applying the style of one image to the content of another. Classically this was a problem reserved for only artists, but more recently this has also interested computer scientists. There are several different ideas on how this can be achieved, ranging from how to separate the style from the content, to how well an algorithm can generalize. An overview of all the different challenges and solutions will be given in this chapter.

2.2.1 Datasets

Due to a lack of benchmark datasets, multiple papers will mix and match from different datasets, like COCO or ImageNet [65].

Cityscape Dataset [66] consists of 2975 images of cityscapes with semantic annotations.

Facades Dataset [67] consists of 400 images of building facades with architectural annotations.

Maps Dataset [74] consists of 1096 images of maps and areal photos gathered from Google Maps around New York City.

Edges2shoes Dataset [?] consists of 50,000 paired images between edges and photos of shoes.

Edges2handbags Dataset [?] consists of 137,000 paired images between edges and photos of handbags.

Season transfer Dataset [60] consists of 2127 images of Yosemite during summer and winter downloaded from Flickr.

Night2Day Dataset [?] consists of 20,000 images taken from time-lapse datasets and annotated through crowdsourcing.

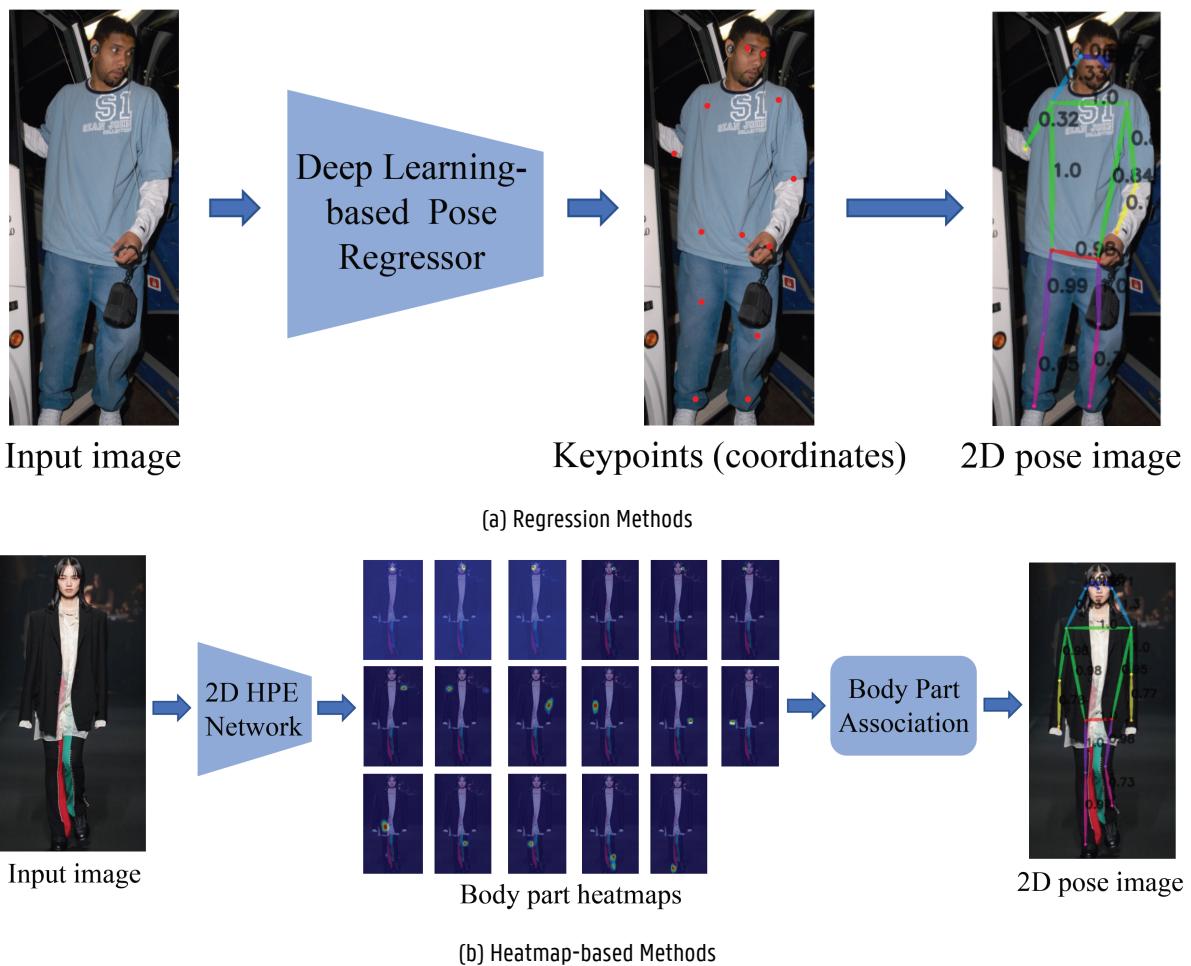


Figure 2.3: The different methods of single-person human pose estimation.[3]

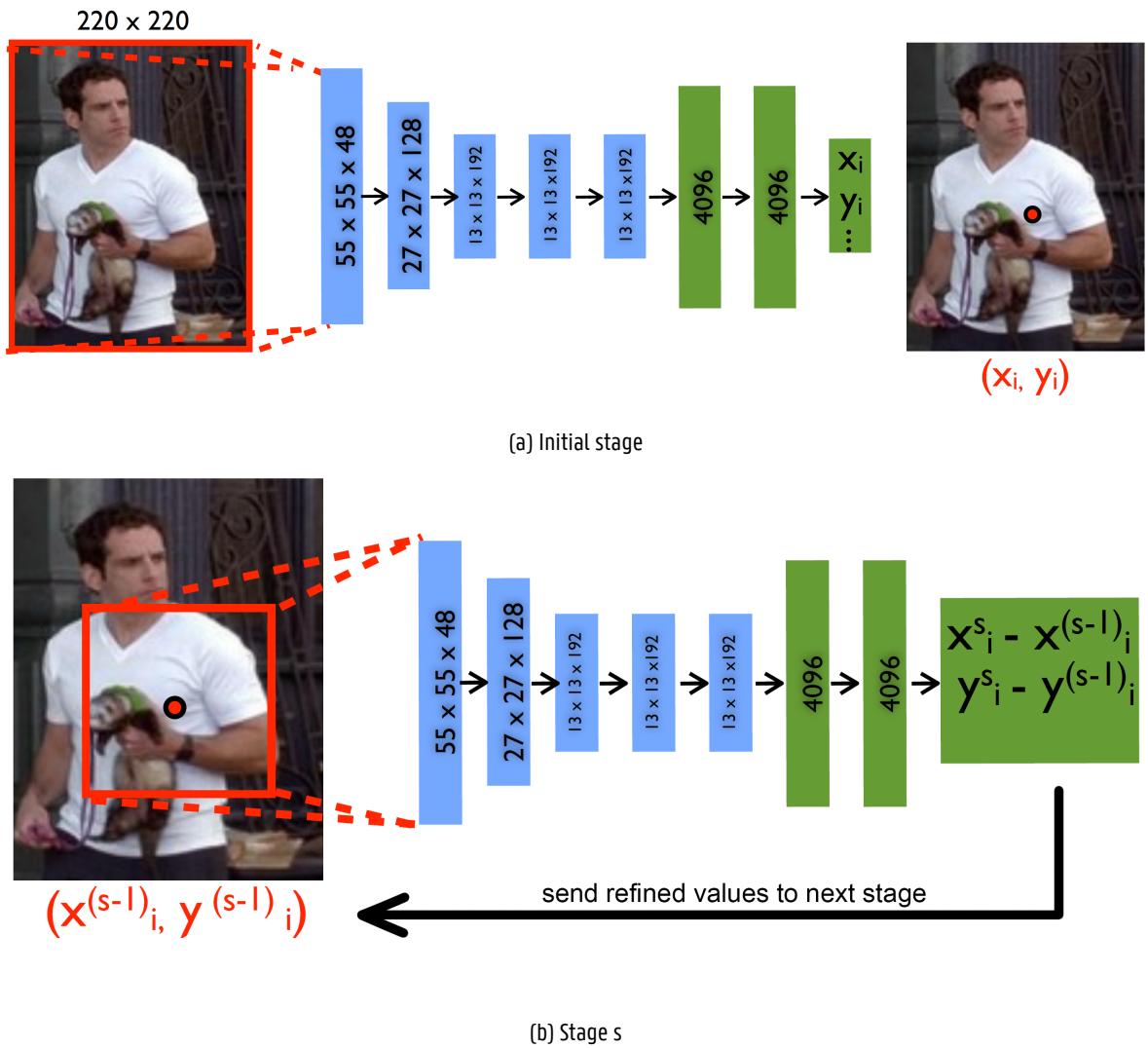


Figure 2.4: Convolution layers in blue and fully connected layers in green. The initial stage is applied to the whole images, while in stage s it will work on a sub-image based on the result of the previous stage.[4]

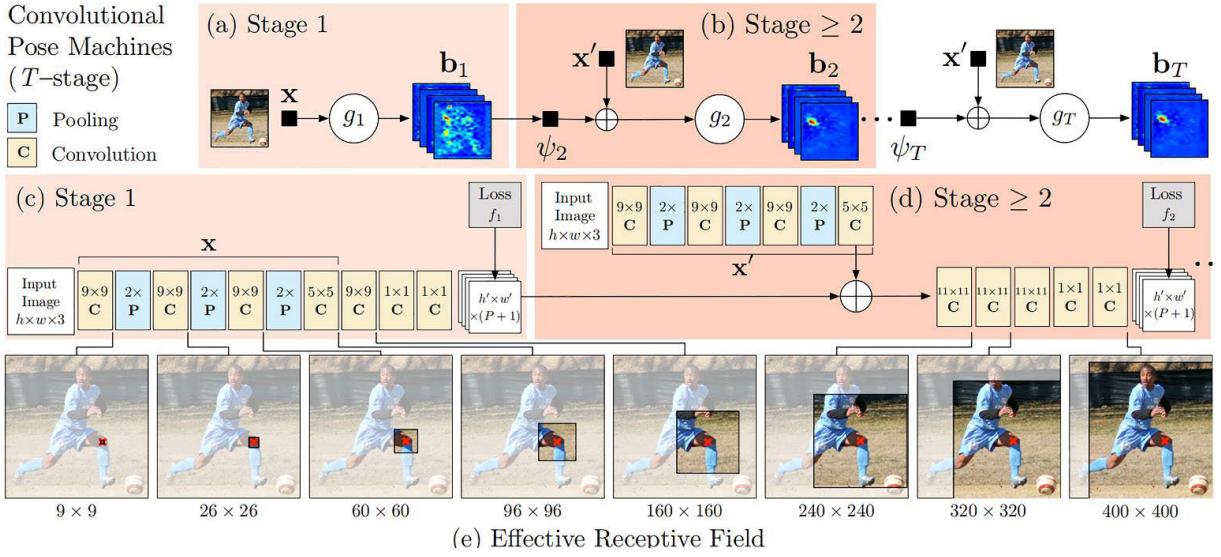


Figure 2.5: Architecture and receptive fields of CPMs. (a) and (b) represent the pose machine architecture.[5] (c) and (d) show the corresponding convolutional networks used by CPMs.[6]

ing.

WikiArt Dataset [68] consists of 80,000 fine-art paintings. All are annotated for 27 styles, 60,000 are annotated for 20 genres and 20,000 for 23 artists.

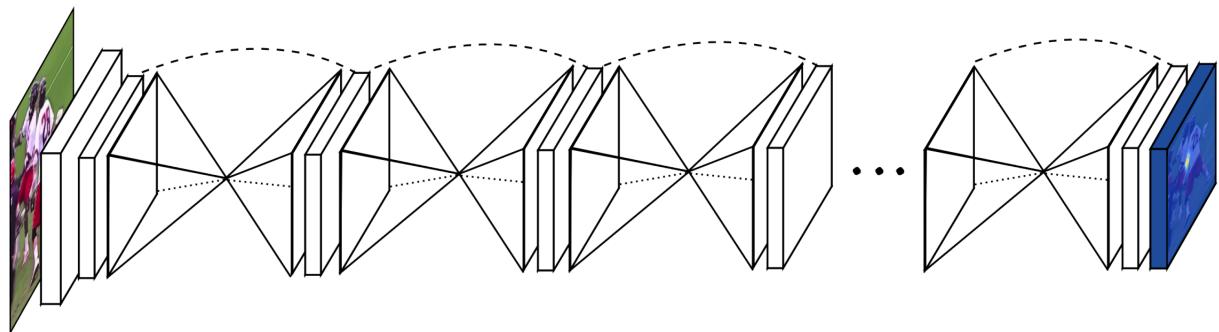
2.2.2 Optimization-based Networks

Gatys et al. [10] introduce deep neural networks to image style transfer. Using a modified VGG-network [69], they extract the features of an image by reconstructing the content from the feature maps in the higher layers on a white noise image. The same is done for the style of the other image. It extracts the style representation of the image by using the Gram matrix to represent style features of the image and then reconstructs it on the same white noise image. The Gram matrix is the vector product of two sets of vectorized feature maps. This method is shown in 2.9. They remark that the resolution of the images affects the performance of the algorithm and is thus restricted to low resolutions. At the same time, the synthesized images contain some low-level noise, but this can possibly be removed with a denoiser.

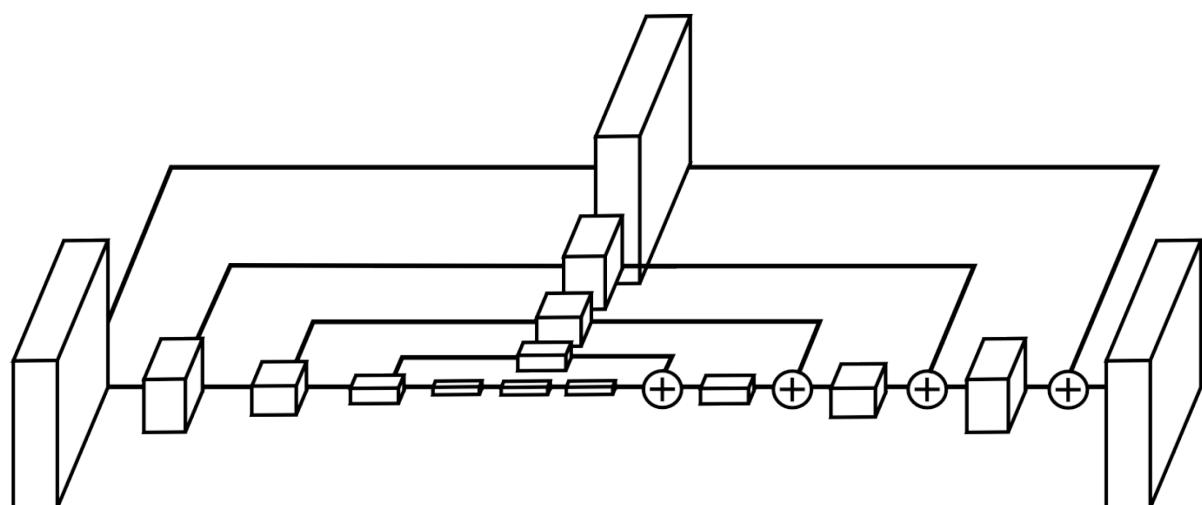
2.2.3 Feed-forward Generation Networks

To improve the performance, Ulyanov et al. [11] suggest the use of a feed-forward generation network instead of back-propagation. Backpropagation requires an iterative process to change the pixel values to match the desired statistics. A feed-forward network can do this in a single evaluation. To train such a network they use a pre-trained network for image classification, and calculate a texture and content loss like [10], as shown in 2.10. Johnson et al. [12] propose a very similar method as can be seen in 2.11.

Since their contribution did increase the speed, but at the expense of quality, Ulyanov et al. [13] suggest further improvements to their network. First, they replace BN [70] with IN which alone has a significant impact on quality as can



(a) Stacked Hourglass



(b) Hourglass Module

Figure 2.6: The structure of a "stacked hourglass" network and a single "hourglass" module.[7]

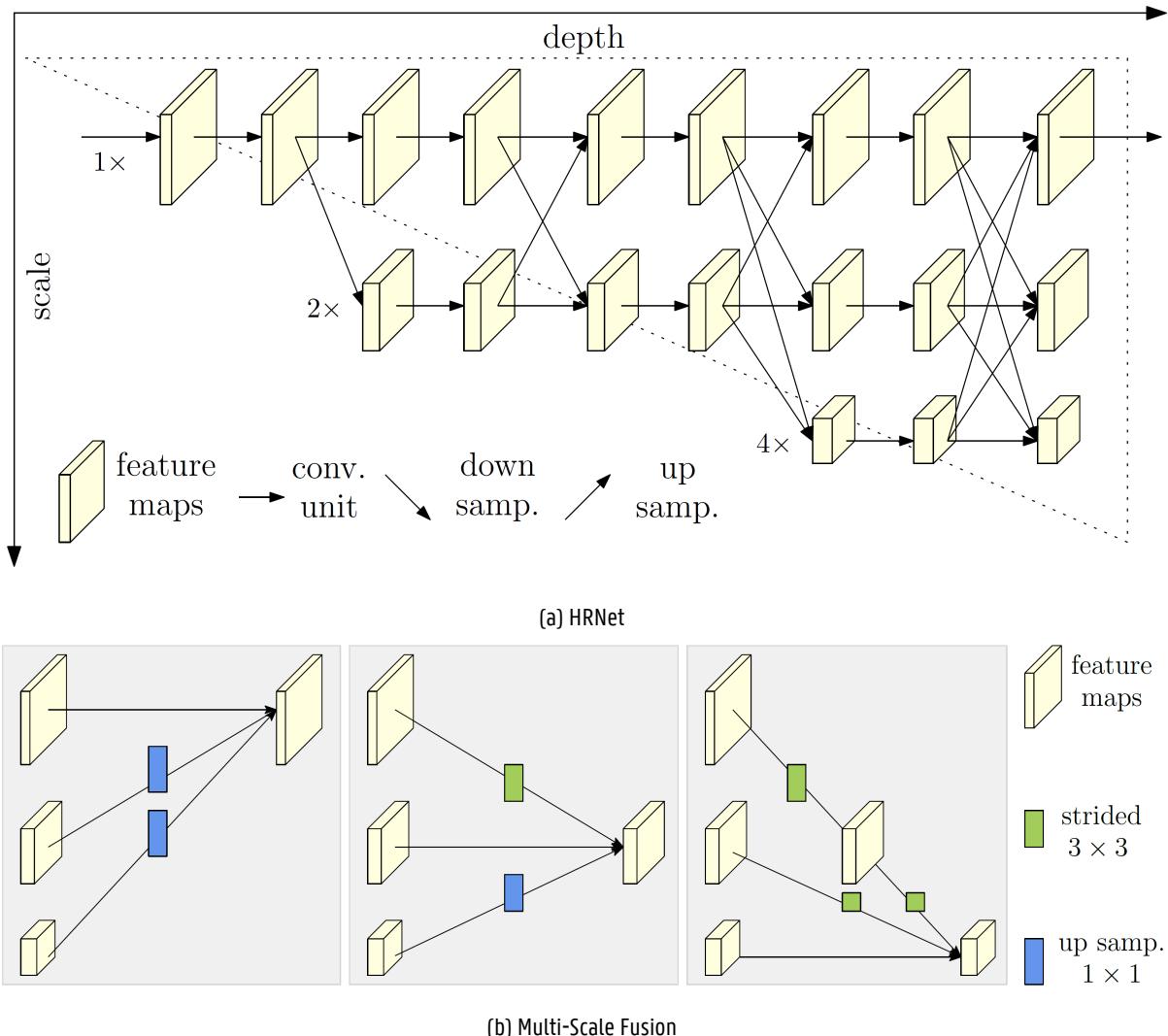


Figure 2.7: The architecture of the High-Resolution network and how it applies multi-scale fusion.[8]

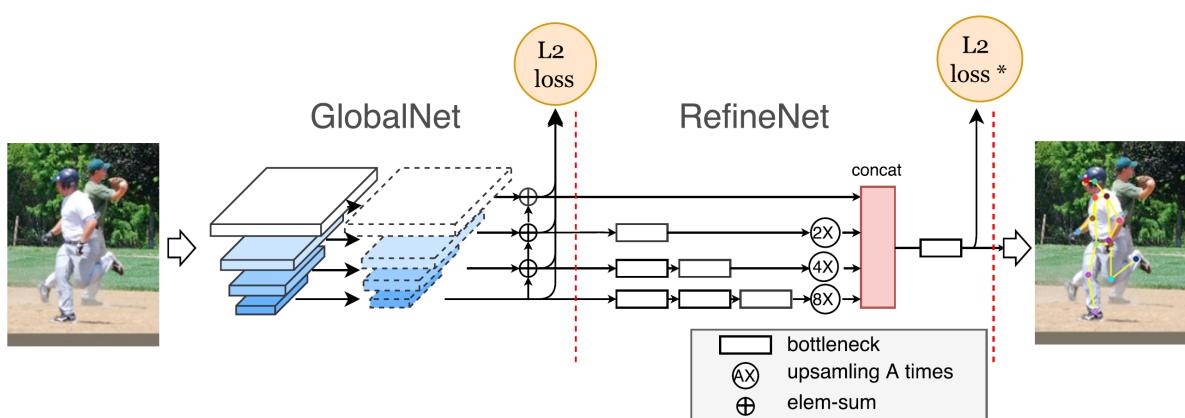


Figure 2.8: Cascaded Pyramid Network. “L2 loss*” means L2 loss with online hard keypoints mining.[9]

be seen in 2.12. Second, they learn the generator to sample from the Julesz ensemble [71] which improves variation in the outputs.

Dumoulin et al. [72] note that previous feed-forward networks are limited to one style. In order to facilitate many different styles, there would need to be a network trained separately for each which limits the applications for mobile devices. In order to make the network more memory efficient, they propose a conditional style transfer network; given a content image and a style name, it transforms the image to the corresponding style. They argue that after normalization each style can be distinguished by specializing scaling and shifting parameters. They call this Conditional Instance Normalization (CIN). Since it only changes the scale and shift parameters for different styles, the network requires fewer parameters. Of the 1.6M parameters, only 3K are needed for the different styles.

Another network that puts a focus on multiple styles comes from Chen et al. [9]. They propose a StyleBank, which can store multiple convolution filter banks each representing a different style. They use an auto-encoder network with in between a StyleBank layer. During training, for each $T + 1$ iterations the entire network is first trained with a perception loss for the first T iterations. Then only the auto-encoder network is trained with a Mean Square Error (MSE) loss. This way the auto-encoder only retains the content and the StyleBank layer only the different styles. This also allows to lock the encoder and decoder to learn a new style afterwards.

While CIN allows for multiple styles, it's still limited to the ones that were seen during training. Huang et al. [73] try to remedy this by introducing an Adaptive Instance Normalization (AdaIN) layer. Unlike the other normalization techniques, AdaIN does not have affine parameters, and will adaptively compute these from the style image. 2.13 shows how well the different networks can handle unseen styles.

2.2.4 Generative Adversarial Networks

With the introduction of GANs, the quality of generative models have greatly increased. It is not surprising then that this got picked up in research for NST.

Among the first was Isola et al. [74] who use a cGAN. With cGAN, the generator network has an extra input which here is the image to be translated. They use the network from [75] which uses modules of the form convolution-BatchNorm-ReLu[70]. Additionally, in order to pass shared features in the generator they add skip connections like with "U-Net" [76]. For the discriminator, which they call PatchGAN, they validate $N \times N$ patches and take the average as output. They take this loss together with the $L1$ loss because $L2$ loss produces blurry results.

This still requires paired training samples, while Taigman et al. [14] are doing research in unsupervised domain transfer. Domain transfer can be used for NST, but this is not possible the other way around. Their network uses a encoder-decoder as the generator and they assume that $f(x)$ is constant between 2 domains. The discriminator has a ternary output and distinguishes between real, fake and reconstruction. They add several new loss functions which check the consistency between the 2 domains (consistency loss) and if G performs perfect reconstruction (reconstruction loss). This can be seen in 2.15. For f , they use a pre-trained network that is trained on paired samples.

In order to make the network completely unsupervised, Yi et al.[77] propose DualGAN, Kim et al. [56] DiscoGAN and Zhu et al. [60] CycleGAN, which are all 3 essentially the same proposal. The entire model consists of 2 cycle-consistent networks where each translates from one domain to the other. A cycle-consistent network will first translate the input to target domain and then back to the original domain. Each domain has a discriminator which compares the real input from one network with the fake from the other; the adversarial loss. As seen in 2.16b. In addition to this there's a cycle-consistency

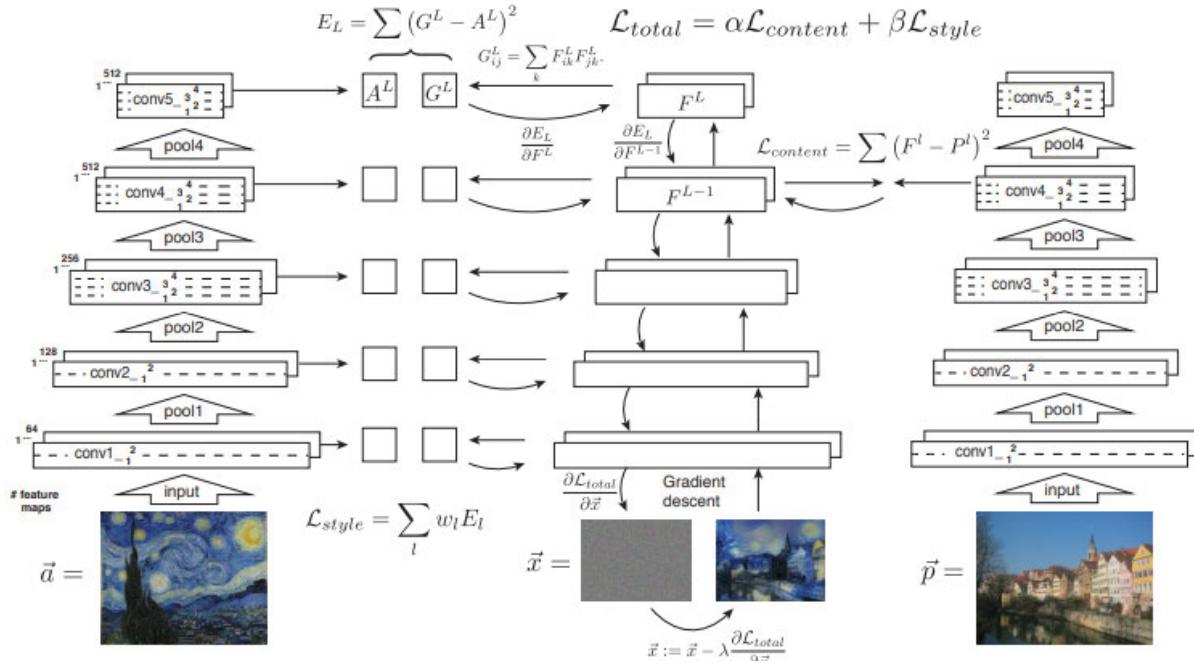


Figure 2.9: Style transfer algorithm. (Gatys et al. [10]).

loss, which is the MSE between the input and the reconstructed image as you can see in 2.16a. The goal is to minimize the adversarial and cycle-consistency losses, while maximizing the discriminators' accuracy. Zhu et al. [60] also introduce an identity loss.

Liu et al. [15] introduce the latent space assumption which assumes that paired images from different domains can be mapped to a shared latent space with the same latent representation. The network consists of 2 domain image encoders E_1 and E_2 , 2 domain image generators G_1 and G_2 , and 2 domain discriminators D_1 and D_2 . As can be seen in 2.17. The encoders and generators are paired and form a Variational Autoencoder (VAE) [78]. The encoder maps the input to latent space, and the generator reconstructs the image. This is the reconstruction loss. They use weight-sharing, which shares the weight of the last 2 layers of the encoders and of the first 2 layers of the generators. The generators and discriminators are paired to form a GAN. The generator can also construct an image from the latent code from the other encoder's input. This image is used to train the GAN. They also show that the shared-latent space assumption implies cycle-consistency, which is the final loss function of the network.

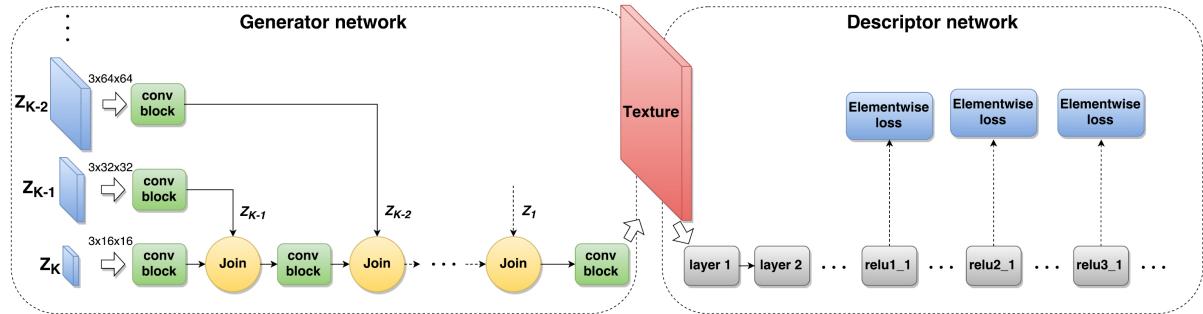


Figure 2.10: A texture network by Ulyanov et al. [11]. The generator network (left) is the only one that changes. The descriptor network (right) is used to calculate the loss.

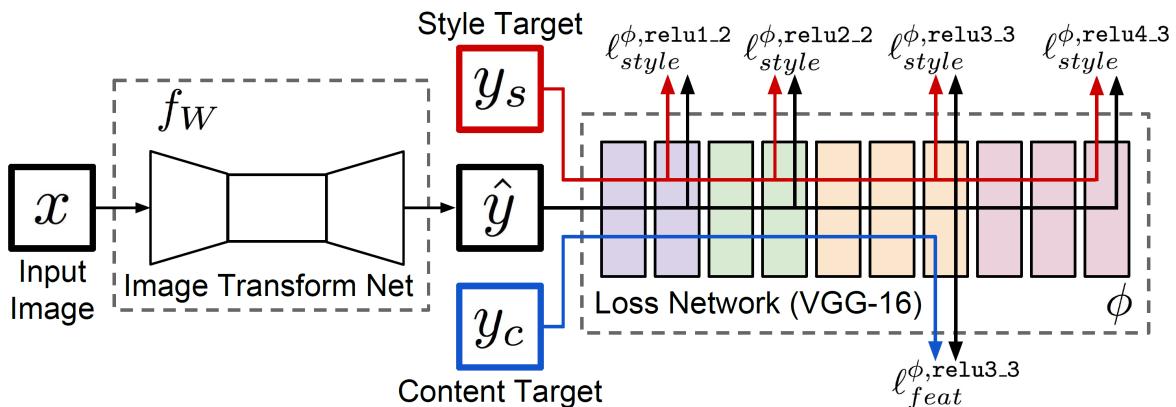
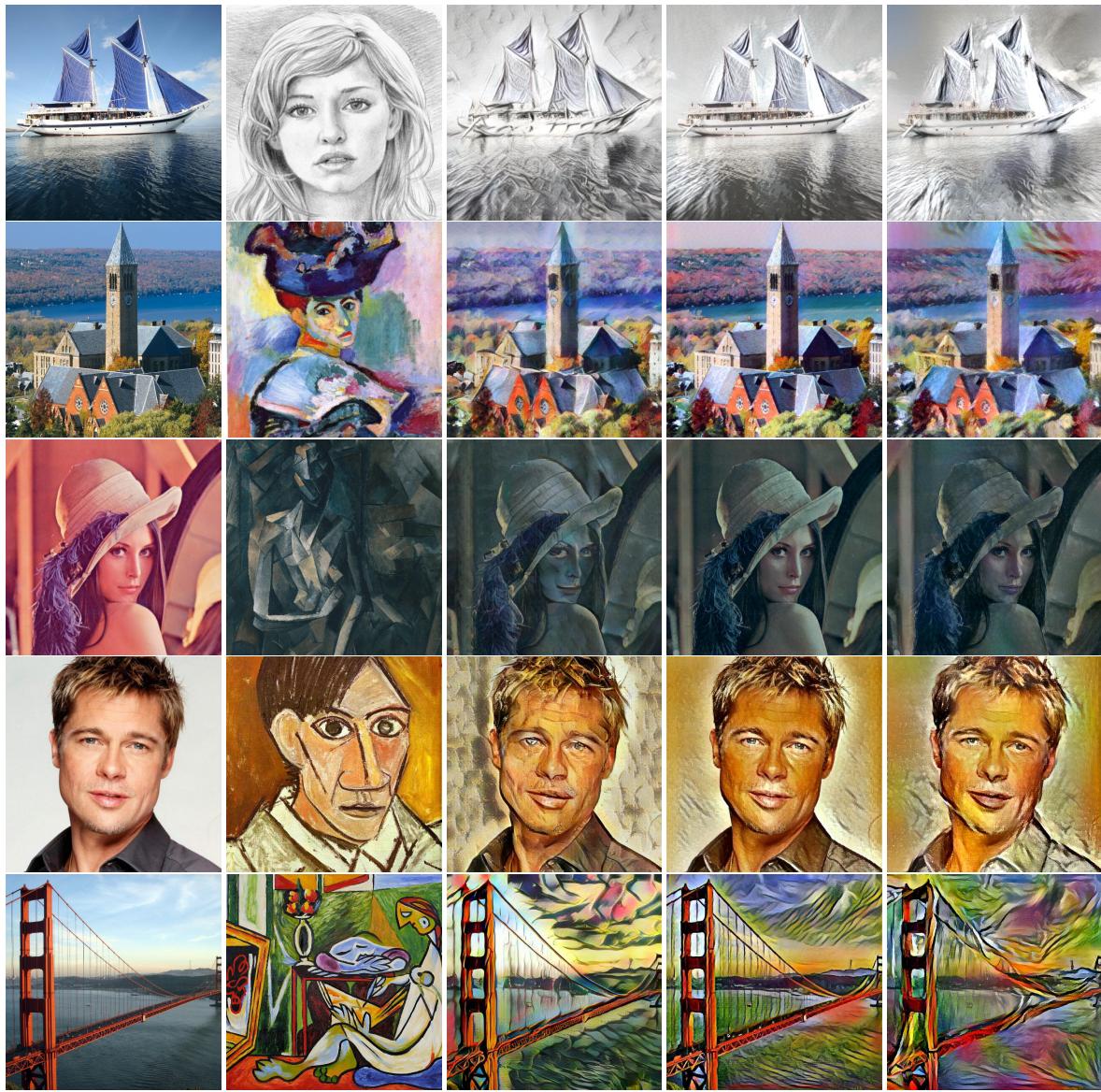


Figure 2.11: An image transformation network by Johnson et al. [12]. The image transform network (left) is the only one that changes. A loss network (right) is used to define perceptual loss functions.



Figure 2.12: A comparison between (c) BN and (d) IN.[13]



(a) Content Image

(b) Style Image

(c) Huang et al.

(d) Ulyanov et al.

(e) Gatys et al.

Figure 2.13: A comparison between different style transfers where the style was not seen during training.

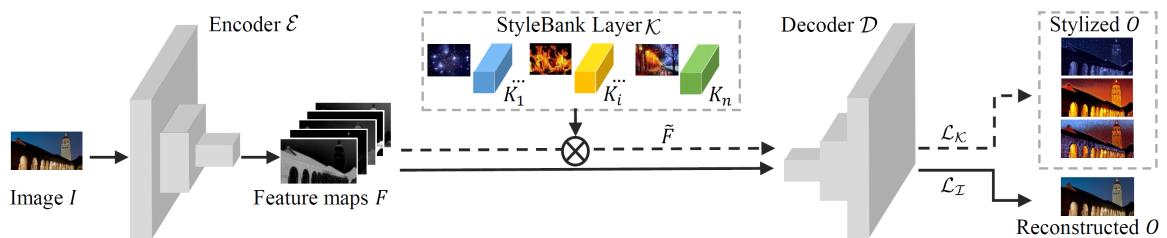


Figure 2.14: The stylebank network by Chen et al. [9].

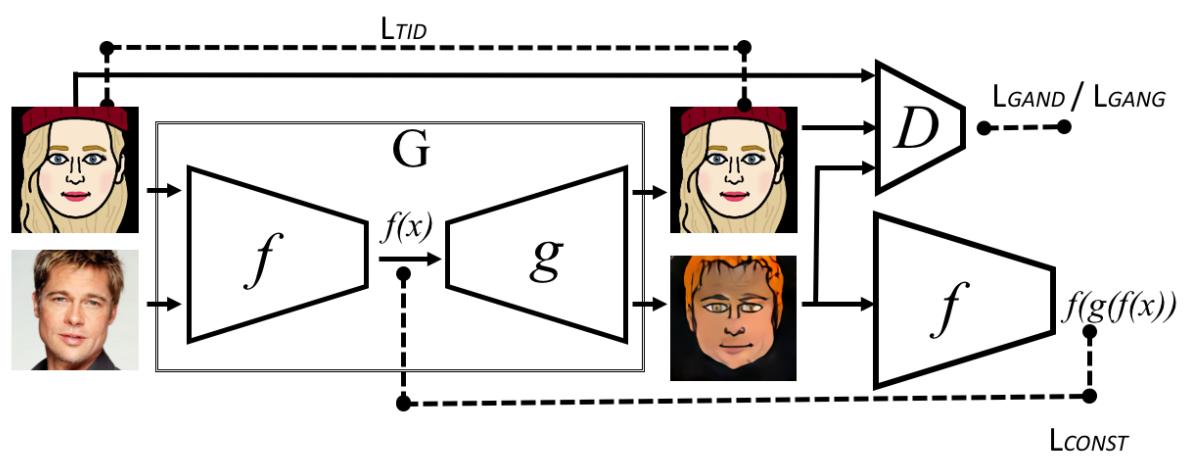
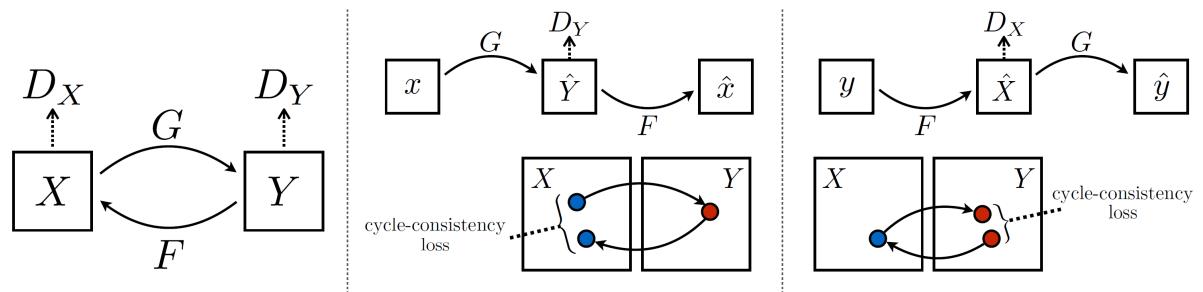
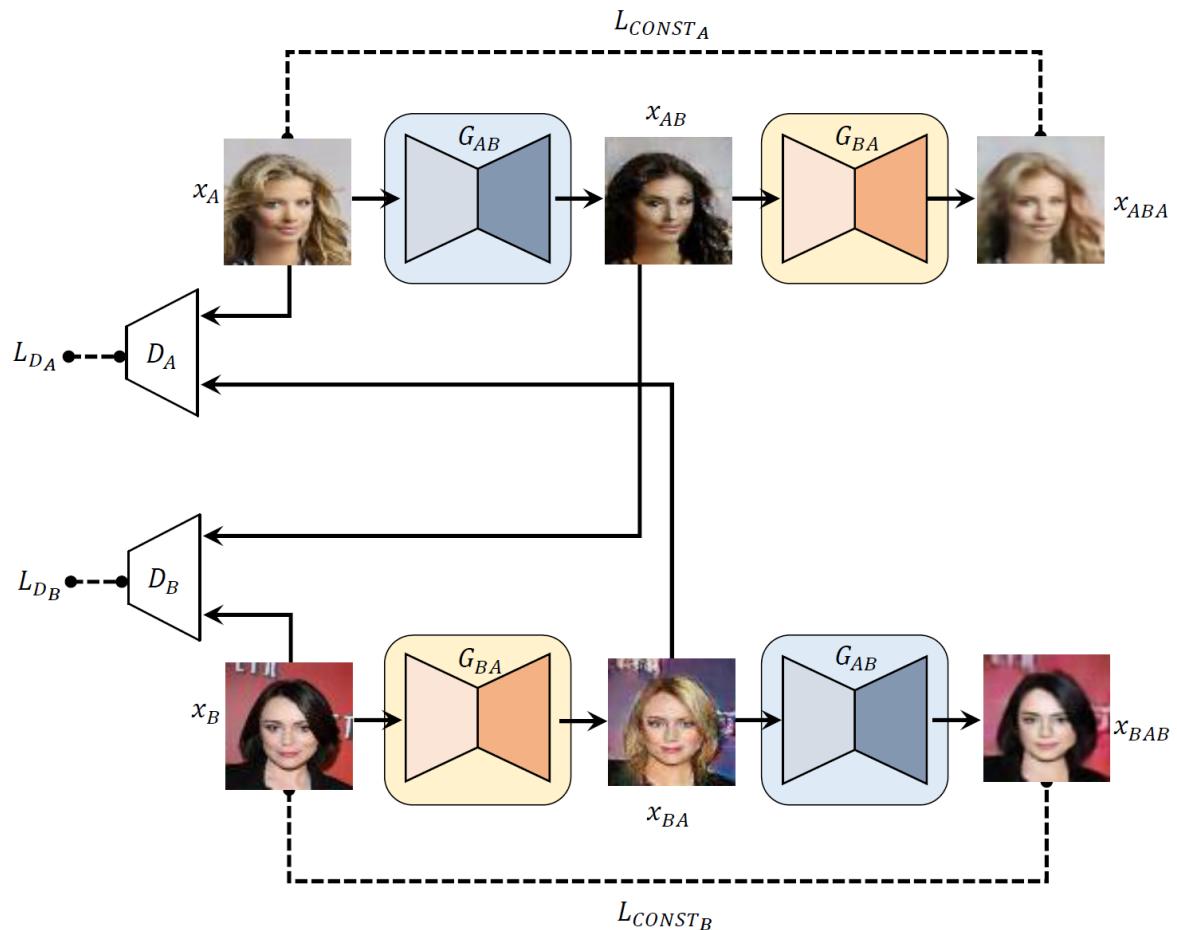


Figure 2.15: The domain transfer network by Taigman et al. [14].



(a) As illustrated by Zhu et al. [60].



(b) As illustrated by Kim et al. [56].

Figure 2.16: The cycle-consistent network.

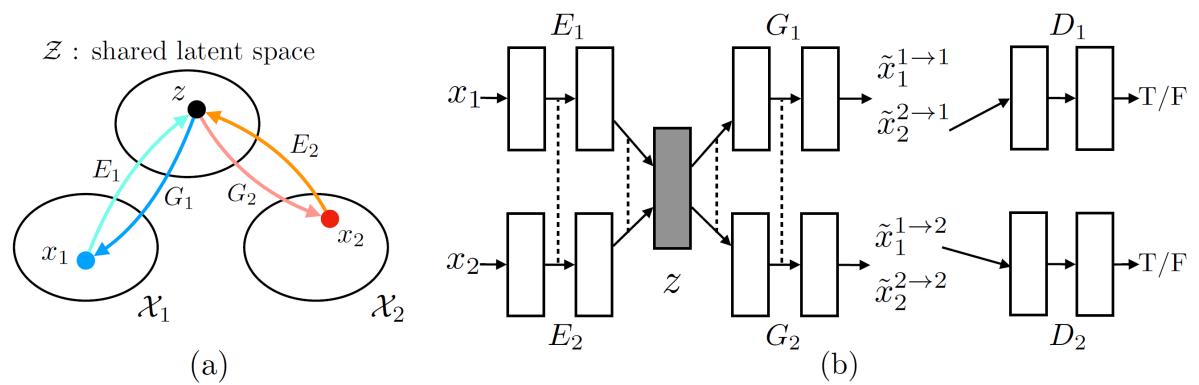


Figure 2.17: (a) The shared latent space assumption. (b) The unsupervised image-to-image translation network by Liu et al. [15].

3

Titel tweede hoofdstuk

Vul aan.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

3.1 Sectie titel

Vul aan.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent

vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor
orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo. Voorbeeld figuur.

4

Titel derde hoofdstuk

Vul aan.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

4.1 Sectie titel

Vul aan.

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.

Sed eleifend, eros sit amet faucibus elementum, urna sapien consectetur mauris, quis egestas leo justo non risus. Morbi non felis ac libero vulputate fringilla. Mauris libero eros, lacinia non, sodales quis, dapibus porttitor, pede. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi dapibus mauris condimentum nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam sit amet erat. Nulla varius. Etiam tincidunt dui vitae turpis. Donec leo. Morbi vulputate convallis est. Integer aliquet. Pellentesque aliquet sodales urna.

Nullam eleifend justo in nisl. In hac habitasse platea dictumst. Morbi nonummy. Aliquam ut felis. In velit leo, dictum vitae, posuere id, vulputate nec, ante. Maecenas vitae pede nec dui dignissim suscipit. Morbi magna. Vestibulum id purus eget velit laoreet laoreet. Praesent sed leo vel nibh convallis blandit. Ut rutrum. Donec nibh. Donec interdum. Fusce sed pede sit amet elit rhoncus ultrices. Nullam at enim vitae pede vehicula iaculis.

4.2 Sectie titel2

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Voorbeeld figuur.

Subtitel

Vul aan

4.2.1 Functie is_isbn

Voorbeeld listing.

```
def is_isbn(isbn: str) -> bool:
    if not type(isbn) is str:
        return False
    if len(isbn) != 10:
        return False
    if not isbn[:9].isdigit():
        return False
    if not isbn[9] in "0123456789X":
        return False
    som = 0
    for i in range(9):
        cijfer = int(isbn[i])
        som += (i + 1) * cijfer
    if isbn[9] == "X":
        laatste_cijfer = 10
    else:
        laatste_cijfer = int(isbn[9])
    return laatste_cijfer == som % 11
```

Code Fragment 4.1: Functie is_isbn

Conclusie

Vul aan.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Ethische en maatschappelijke reflectie

Vul aan.

Meer informatie kan je opzoeken op <https://www.sdgs.be/nl/sdgs>

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Referenties

- [1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3686–3693.
- [2] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *CoRR*, vol. abs/2006.01423, 2020. [Online]. Available: <https://arxiv.org/abs/2006.01423>
- [3] C. Zheng, W. Wu, T. Yang, S. Zhu, C. Chen, R. Liu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," *CoRR*, vol. abs/2012.13392, 2020. [Online]. Available: <https://arxiv.org/abs/2012.13392>
- [4] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2014. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2014.214>
- [5] V. Ramakrishna, D. Munoz, M. Hebert, J. Andrew Bagnell, and Y. Sheikh, "Pose machines: Articulated pose estimation via inference machines," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 33–47.
- [6] S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," *CoRR*, vol. abs/1602.00134, 2016. [Online]. Available: <http://arxiv.org/abs/1602.00134>
- [7] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," *CoRR*, vol. abs/1603.06937, 2016. [Online]. Available: <http://arxiv.org/abs/1603.06937>
- [8] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," *CoRR*, vol. abs/1902.09212, 2019. [Online]. Available: <http://arxiv.org/abs/1902.09212>
- [9] Y. Chen, C. Shen, X. Wei, L. Liu, and J. Yang, "Adversarial posenet: A structure-aware convolutional network for human pose estimation," *CoRR*, vol. abs/1705.00389, 2017. [Online]. Available: <http://arxiv.org/abs/1705.00389>
- [10] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, 2016. [Online]. Available: <https://api.semanticscholar.org/CorpusID:206593710>
- [11] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," *CoRR*, vol. abs/1603.03417, 2016. [Online]. Available: <http://arxiv.org/abs/1603.03417>
- [12] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016. [Online]. Available: <http://arxiv.org/abs/1603.08155>
- [13] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4105–4113.
- [14] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," *CoRR*, vol. abs/1611.02200, 2016. [Online]. Available: <http://arxiv.org/abs/1611.02200>

- [15] M. Liu, T. M. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *CoRR*, vol. abs/1703.00848, 2017. [Online]. Available: <http://arxiv.org/abs/1703.00848>
- [16] M.-C. Marinescu, A. Reshetnikov, and J. M. López, "Improving object detection in paintings based on time contexts," in *2020 International Conference on Data Mining Workshops (ICDMW)*, 2020, pp. 926–932.
- [17] M. Sabatelli, N. Banar, M. Cocriamont, E. Coudyzer, K. Lasaracina, W. Daelemans, P. Geurts, and M. Kestemont, "Advances in digital music iconography: Benchmarking the detection of musical instruments in unrestricted, non-photorealistic images from the artistic domain," *Digital Humanities Quarterly*, vol. 15, no. 1, February 2021.
- [18] T. L. Munea, Y. Z. Jembre, H. T. Weldegeebriel, L. Chen, C. Huang, and C. Yang, "The progress of human pose estimation: A survey and taxonomy of models applied in 2d human pose estimation," *IEEE Access*, vol. 8, pp. 133 330–133 348, 2020.
- [19] W. Liu, Q. Bao, Y. Sun, and T. Mei, "Recent advances in monocular 2d and 3d human pose estimation: A deep learning perspective," *CoRR*, vol. abs/2104.11536, 2021. [Online]. Available: <https://arxiv.org/abs/2104.11536>
- [20] H. Chen, R. Feng, S. Wu, H. Xu, F. Zhou, and Z. Liu, "2d human pose estimation: a survey," *Multimedia Systems*, pp. 1–24, 2022.
- [21] A. Jain, J. Tompson, M. Andriluka, G. W. Taylor, and C. Bregler, "Learning human pose estimation features with convolutional networks," 2014.
- [22] Z. Luo, Z. Wang, Y. Huang, T. Tan, and E. Zhou, "Rethinking the heatmap regression for bottom-up human pose estimation," *CoRR*, vol. abs/2012.15175, 2020. [Online]. Available: <https://arxiv.org/abs/2012.15175>
- [23] S. Ju, M. Black, and Y. Yacoob, "Cardboard people: a parameterized model of articulated image motion," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 38–44.
- [24] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1077314285710041>
- [25] H. Sidenbladh, F. De la Torre, and M. Black, "A framework for modeling the appearance of 3d articulated figures," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 2000, pp. 368–375.
- [26] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. Black, "Smpl: a skinned multi-person linear model," vol. 34, 11 2015.
- [27] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," in *British Machine Vision Conference*, 2010. [Online]. Available: <https://api.semanticscholar.org/CorpusID:7318714>
- [28] ——, "Learning effective human pose estimation from inaccurate annotation," in *CVPR 2011*, 2011, pp. 1465–1472.
- [29] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: <http://arxiv.org/abs/1405.0312>

- [30] B. Sapp and B. Taskar, "Modec: Multimodal decomposable models for human pose estimation," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3674–3681.
- [31] J. Li, C. Wang, H. Zhu, Y. Mao, H. Fang, and C. Lu, "Crowdpose: Efficient crowded scenes pose estimation and A new benchmark," *CoRR*, vol. abs/1812.00324, 2018. [Online]. Available: <http://arxiv.org/abs/1812.00324>
- [32] G. Pons-Moll and B. Rosenhahn, *Model-Based Pose Estimation*. London: Springer London, 2011, pp. 139–170. [Online]. Available: https://doi.org/10.1007/978-0-85729-997-0_9
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [34] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," *CoRR*, vol. abs/1507.06550, 2015. [Online]. Available: <http://arxiv.org/abs/1507.06550>
- [35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: <http://arxiv.org/abs/1409.4842>
- [36] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," *CoRR*, vol. abs/1406.2984, 2014. [Online]. Available: <http://arxiv.org/abs/1406.2984>
- [37] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," *CoRR*, vol. abs/1411.4280, 2014. [Online]. Available: <http://arxiv.org/abs/1411.4280>
- [38] W. Yang, S. Li, W. Ouyang, H. Li, and X. Wang, "Learning feature pyramids for human pose estimation," *CoRR*, vol. abs/1708.01101, 2017. [Online]. Available: <http://arxiv.org/abs/1708.01101>
- [39] C. Chou, J. Chien, and H. Chen, "Self adversarial training for human pose estimation," *CoRR*, vol. abs/1707.02439, 2017. [Online]. Available: <http://arxiv.org/abs/1707.02439>
- [40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [41] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, "Bottom-up higher-resolution networks for multi-person pose estimation," *CoRR*, vol. abs/1908.10357, 2019. [Online]. Available: <http://arxiv.org/abs/1908.10357>
- [42] C. Yu, B. Xiao, C. Gao, L. Yuan, L. Zhang, N. Sang, and J. Wang, "Lite-hrnet: A lightweight high-resolution network," *CoRR*, vol. abs/2104.06403, 2021. [Online]. Available: <https://arxiv.org/abs/2104.06403>
- [43] Y. Yuan, R. Fu, L. Huang, W. Lin, C. Zhang, X. Chen, and J. Wang, "Hrformer: High-resolution transformer for dense prediction," *CoRR*, vol. abs/2110.09408, 2021. [Online]. Available: <https://arxiv.org/abs/2110.09408>
- [44] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

- [45] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [46] C. Chou, J. Chien, and H. Chen, "Self adversarial training for human pose estimation," *CoRR*, vol. abs/1707.02439, 2017. [Online]. Available: <http://arxiv.org/abs/1707.02439>
- [47] U. Iqbal and J. Gall, "Multi-person pose estimation with local joint-to-person associations," *CoRR*, vol. abs/1608.08526, 2016. [Online]. Available: <http://arxiv.org/abs/1608.08526>
- [48] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: <http://arxiv.org/abs/1506.01497>
- [49] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, and B. Schiele, "Deepcut: Joint subset partition and labeling for multi person pose estimation," *CoRR*, vol. abs/1511.06645, 2015. [Online]. Available: <http://arxiv.org/abs/1511.06645>
- [50] H. Fang, S. Xie, and C. Lu, "RMPE: regional multi-person pose estimation," *CoRR*, vol. abs/1612.00137, 2016. [Online]. Available: <http://arxiv.org/abs/1612.00137>
- [51] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. P. Murphy, "Towards accurate multi-person pose estimation in the wild," *CoRR*, vol. abs/1701.01779, 2017. [Online]. Available: <http://arxiv.org/abs/1701.01779>
- [52] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," *CoRR*, vol. abs/1612.03144, 2016. [Online]. Available: <http://arxiv.org/abs/1612.03144>
- [53] T. Kong, A. Yao, Y. Chen, and F. Sun, "Hypernet: Towards accurate region proposal generation and joint object detection," *CoRR*, vol. abs/1604.00600, 2016. [Online]. Available: <http://arxiv.org/abs/1604.00600>
- [54] K. Su, D. Yu, Z. Xu, X. Geng, and C. Wang, "Multi-person pose estimation with enhanced channel-wise and spatial information," *CoRR*, vol. abs/1905.03466, 2019. [Online]. Available: <http://arxiv.org/abs/1905.03466>
- [55] W. Li, Z. Wang, B. Yin, Q. Peng, Y. Du, T. Xiao, G. Yu, H. Lu, Y. Wei, and J. Sun, "Rethinking on multi-stage networks for human pose estimation," *CoRR*, vol. abs/1901.00148, 2019. [Online]. Available: <http://arxiv.org/abs/1901.00148>
- [56] Y. Kim, C. Denton, L. Hoang, and A. M. Rush, "Structured attention networks," *CoRR*, vol. abs/1702.00887, 2017. [Online]. Available: <http://arxiv.org/abs/1702.00887>
- [57] S. Yang, Z. Quan, M. Nie, and W. Yang, "Transpose: Towards explainable human pose estimation by transformer," *CoRR*, vol. abs/2012.14214, 2020. [Online]. Available: <https://arxiv.org/abs/2012.14214>
- [58] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepcut: A deeper, stronger, and faster multi-person pose estimation model," *CoRR*, vol. abs/1605.03170, 2016. [Online]. Available: <http://arxiv.org/abs/1605.03170>
- [59] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *CoRR*, vol. abs/1812.08008, 2018. [Online]. Available: <http://arxiv.org/abs/1812.08008>

- [60] X. Zhu and Y. Jiang, "Multi-person pose estimation for posetrack with enhanced part affinity fields," 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:52563463>
- [61] G. Hidalgo, Y. Raaj, H. Idrees, D. Xiang, H. Joo, T. Simon, and Y. Sheikh, "Single-network whole-body pose estimation," *CoRR*, vol. abs/1909.13423, 2019. [Online]. Available: <http://arxiv.org/abs/1909.13423>
- [62] S. Kreiss, L. Bertoni, and A. Alahi, "Pifpaf: Composite fields for human pose estimation," *CoRR*, vol. abs/1903.06593, 2019. [Online]. Available: <http://arxiv.org/abs/1903.06593>
- [63] A. Newell and J. Deng, "Associative embedding: End-to-end learning for joint detection and grouping," *CoRR*, vol. abs/1611.05424, 2016. [Online]. Available: <http://arxiv.org/abs/1611.05424>
- [64] Z. Luo, Z. Wang, Y. Huang, T. Tan, and E. Zhou, "Rethinking the heatmap regression for bottom-up human pose estimation," *CoRR*, vol. abs/2012.15175, 2020. [Online]. Available: <https://arxiv.org/abs/2012.15175>
- [65] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.
- [66] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," *CoRR*, vol. abs/1604.01685, 2016. [Online]. Available: <http://arxiv.org/abs/1604.01685>
- [67] R. Tylecek and R. Sára, "Spatial pattern templates for recognition of objects with regular structure," in *German Conference on Pattern Recognition*, 2013. [Online]. Available: <https://api.semanticscholar.org/CorpusID:6060524>
- [68] W. R. Tan, C. S. Chan, H. Aguirre, and K. Tanaka, "Improved artgan for conditional synthesis of natural image and artwork," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 394–409, 2019. [Online]. Available: <https://doi.org/10.1109/TIP.2018.2866698>
- [69] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
- [70] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [71] S.-C. Zhu, X. Liu, and Y. N. Wu, "Exploring texture ensembles by efficient markov chain monte carlo-toward a 'trichromacy' theory of texture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 554–569, 2000. [Online]. Available: <https://api.semanticscholar.org/CorpusID:3194236>
- [72] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," *CoRR*, vol. abs/1610.07629, 2016. [Online]. Available: <http://arxiv.org/abs/1610.07629>
- [73] X. Huang and S. J. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," *CoRR*, vol. abs/1703.06868, 2017. [Online]. Available: <http://arxiv.org/abs/1703.06868>
- [74] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016. [Online]. Available: <http://arxiv.org/abs/1611.07004>

- [75] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016.
- [76] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [77] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," *CoRR*, vol. abs/1704.02510, 2017. [Online]. Available: <http://arxiv.org/abs/1704.02510>
- [78] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2022.

Bijlagen

Bijlage 1

Toelichting bijlage.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Bijlage 2

Toelichting bijlage.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus duis, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a,

augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, malesuada id, gravida nec, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, porta sed, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed vitae sem. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Maecenas ante. Duis ullamcorper enim. Donec tristique enim eu leo. Nullam molestie elit eu dolor. Nullam bibendum, turpis vitae tristique gravida, quam sapien tempor lectus, quis pretium tellus purus ac quam. Nulla facilisi.

Duis aliquet dui in est. Donec eget est. Nunc lectus odio, varius at, fermentum in, accumsan non, enim. Aliquam erat volutpat. Proin sit amet nulla ut eros consectetur cursus. Phasellus dapibus aliquam justo. Nunc laoreet. Donec consequat placerat magna. Duis pretium tincidunt justo. Sed sollicitudin vestibulum quam. Nam quis ligula. Vivamus at metus. Etiam imperdiet imperdiet pede. Aenean turpis. Fusce augue velit, scelerisque sollicitudin, dictum vitae, tempor et, pede. Donec wisi sapien, feugiat in, fermentum ut, sollicitudin adipiscing, metus.

Donec vel nibh ut felis consectetur laoreet. Donec pede. Sed id quam id wisi laoreet suscipit. Nulla lectus dolor, aliquam ac, fringilla eget, mollis ut, orci. In pellentesque justo in ligula. Maecenas turpis. Donec eleifend leo at felis tincidunt consequat. Aenean turpis metus, malesuada sed, condimentum sit amet, auctor a, wisi. Pellentesque sapien elit, bibendum ac, posuere et, congue eu, felis. Vestibulum mattis libero quis metus scelerisque ultrices. Sed purus.

Donec molestie, magna ut luctus ultrices, tellus arcu nonummy velit, sit amet pulvinar elit justo et mauris. In pede. Maecenas euismod elit eu erat. Aliquam augue wisi, facilisis congue, suscipit in, adipiscing et, ante. In justo. Cras lobortis neque ac ipsum. Nunc fermentum massa at ante. Donec orci tortor, egestas sit amet, ultrices eget, venenatis eget, mi. Maecenas vehicula leo semper est. Mauris vel metus. Aliquam erat volutpat. In rhoncus sapien ac tellus. Pellentesque ligula.