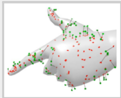
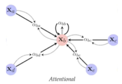
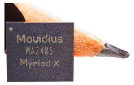
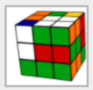
DEEP VISION CONSULTING

2022 INTERNSHIP TOPICS		THEORY 📝 PRACTICE 🔨 COMPLEXITY 🚀	WHAT YOU WILL LEARN
Representations and Architectures for **Geometric Deep Learning**		📝📝📝 🔨🔨 🚀🚀	☐ Theory of geometric deep learning ☐ Apply deep learning to structured data beyond images ☐ Implement and verify theory on real challenging problems
Deep Learning Approaches to **Feature Tracking, Stereo Matching and Optical Flow**		📝 🔨🔨🔨 🚀🚀	☐ Modern DL-based solutions to these classical problems ☐ Secrets of correlation networks ☐ Implement one or more papers and test it on real data
General Strategies for **Self-Supervised Representation Learning**		📝📝📝 🔨 🚀🚀	☐ When and how to exploit the power of self-supervision ☐ The technical and theoretical details behind recent advances ☐ How to implement one or more method and make it work
3D Morphable Models in the Era of Deep Learning		📝📝 🔨🔨🔨 🚀	☐ What DL can do for parametric fitting of complex models ☐ The pros and cons of different 2D-3D methods ☐ How to set up and solve a new fitting problem from scratch
Diffusion Models: what's in it for the industry?		📝📝📝 🔨🔨🔨 🚀🚀🚀	☐ How diffusion models work down to the details ☐ To train a diffusion model on custom dataset and generate from it ☐ Mix cutting edge research material with industry needs
Attacking **NP-HARD problems** with Graph NN		📝📝📝 🔨🔨 🚀🚀🚀	☐ Basics of computational complexity and the main classes of problems ☐ Secrets of graph neural network ☐ How to exploit deep learning beyond classical vision tasks
Deep Learning Inference Engines the **Right Tool for Model Deployment**		📝 🔨🔨🔨 🚀🚀	☐ How to go from training to production of a DL model ☐ Hardware and tools improve speed and portability ☐ Deploy custom DL to microcontrollers and to the cloud
Deep Learning at the Edge: **AI Accelerators**		📝 🔨🔨🔨 🚀🚀	☐ Challenges of running DL on resource-constrained devices ☐ To work with embedded platforms (FPGA, NPU, ARM, …) ☐ Study and implement advanced techniques to improve results
Multi-Modal Deep Learning with an Application to Lip Reading		📝📝 🔨🔨🔨 🚀	☐ How can a DL model work with data from different domains ☐ The scientific method behind building a DL project from zero ☐ Exploit YouTube videos to enrich the training dataset
Beyond OpenCV: Nvidia Npp and Intel Ipp		📝 🔨🔨🔨 🚀🚀	☐ Limits of OpenCV and what can be done to overcome them ☐ Peculiarities of multimedia libraries from Intel and Nvidia ☐ How to exploit them to write efficient image/video operators
Jigsaw Puzzle / Rubik Cube a Combinatorial Problem Solver		📝 🔨🔨🔨 🚀	☐ Formulate the image processing steps required to fetch the input ☐ Understand and code the solution to classical combinatorial problems ☐ Build a computer vision app for mobile that runs in real time

find all the details at www.deepvisionconsulting.com/internships - to apply drop an email at careers@deepvisionconsulting.com

REPRESENTATIONS AND ARCHITECTURES FOR GEOMETRIC DEEP LEARNING

Theory: 📝📝📝 Practice: 🔨🔨 Complexity: 🚀🚀

Pointnet, occupancy networks, graph neural networks, transformers, 3D conv on voxelized inputs or 2D conv that work on depth maps are just a few examples of the way deep learning can be applied to 3D data. The list is far from finished since they are all part of a hot topic - now known as geometric deep learning (https://arxiv.org/abs/2104.13478) - that is rapidly expanding, even as you are reading this.
The aim of this thesis is to draw a panoramic view of the most promising approaches, understanding their weaknesses and strengths, and which assumptions they require to work in practice. You will learn about the theory of geometric deep learning and how to apply deep learning successfully to computer vision data beyond images. A strong candidate will also have the chance to implement and test some of these methods on challenging problems to verify the theoretical conclusions drawn in the first part of the thesis.
The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.

Geometric Deep Learning Blueprint

Let Ω and Ω' be domains, $\mathfrak{G}$ a symmetry group over Ω, and write $\Omega' \subseteq \Omega$ if Ω' can be considered a compact version of Ω.

We define the following building blocks:

Linear $\mathfrak{G}$-equivariant layer $B : \mathcal{X}(\Omega, \mathcal{C}) \rightarrow \mathcal{X}(\Omega', \mathcal{C}')$ satisfying $B(\mathfrak{g}.x) = \mathfrak{g}.B(x)$ for all $\mathfrak{g} \in \mathfrak{G}$ and $x \in \mathcal{X}(\Omega, \mathcal{C})$.

Nonlinearity $\sigma : \mathcal{C} \rightarrow \mathcal{C}'$ applied element-wise as $(\boldsymbol{\sigma}(x))(u) = \sigma(x(u))$.

Local pooling (coarsening) $P : \mathcal{X}(\Omega, \mathcal{C}) \rightarrow \mathcal{X}(\Omega', \mathcal{C})$, such that $\Omega' \subseteq \Omega$.

$\mathfrak{G}$-invariant layer (global pooling) $A : \mathcal{X}(\Omega, \mathcal{C}) \rightarrow \mathcal{Y}$ satisfying $A(\mathfrak{g}.x) = A(x)$ for all $\mathfrak{g} \in \mathfrak{G}$ and $x \in \mathcal{X}(\Omega, \mathcal{C})$.

Using these b
$\mathcal{X}(\Omega, \mathcal{C}) \rightarrow \mathcal{Y}$ o

where the bloc
matches the in
different choic

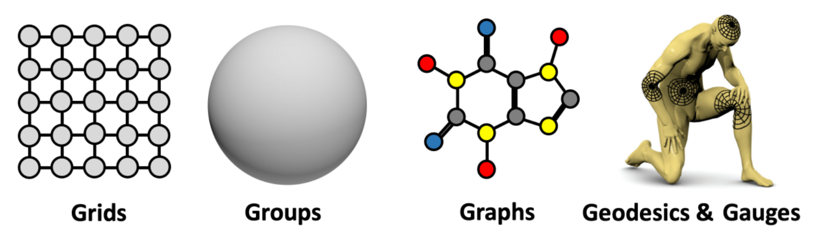


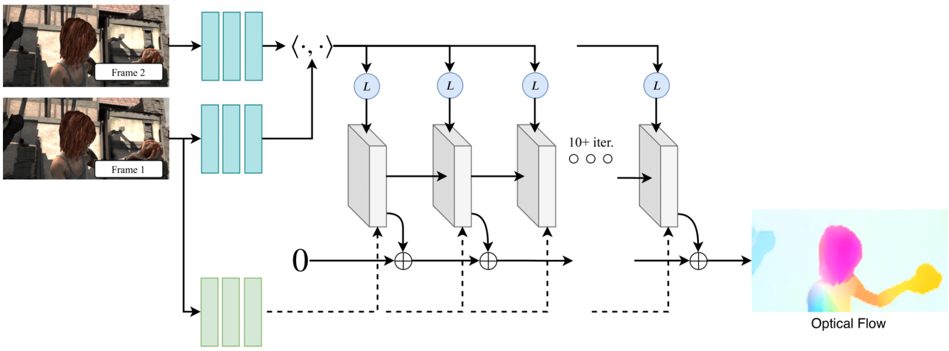
Figure 9: The 5G of Geometric Deep Learning: grids, groups & homogeneous spaces with global symmetry, graphs, geodesics & metrics on manifolds, and gauges (frames for tangent or feature spaces).

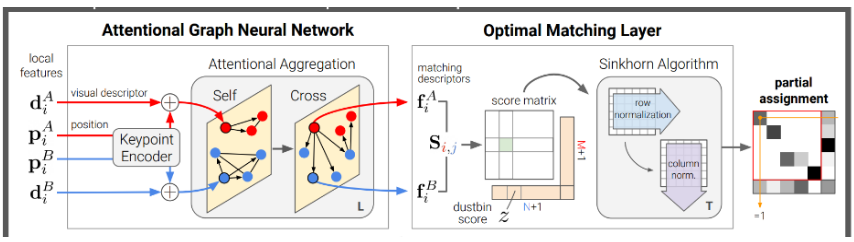
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

MODERN APPROACHES TO FEATURE TRACKING, STEREO MATCHING AND OPTICAL FLOW

Theory: 📝 Practice: 🔨🔨🔨 Complexity: 🚀🚀

Feature tracking, stereo matching and optical flow are three monumental tasks as old as computer vision itself, the last one going back to the early ‘80s. What these tasks have in common is that they must reason across two images to find pixel correspondences, if any. Despite their wide application in all areas of our field and their early development, these techniques have been incredibly resistant to improvements in the last 30 years. Of course, things are now changing with the power of deep learning.
The aim of this thesis is to study these tasks in a unifying perspective and scout the related most interesting research achievements of the deep learning era and, trust us, there are a lot. A strong candidate will also have the chance to implement and test some of these methods on real world problems from the company proprietary datasets. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.





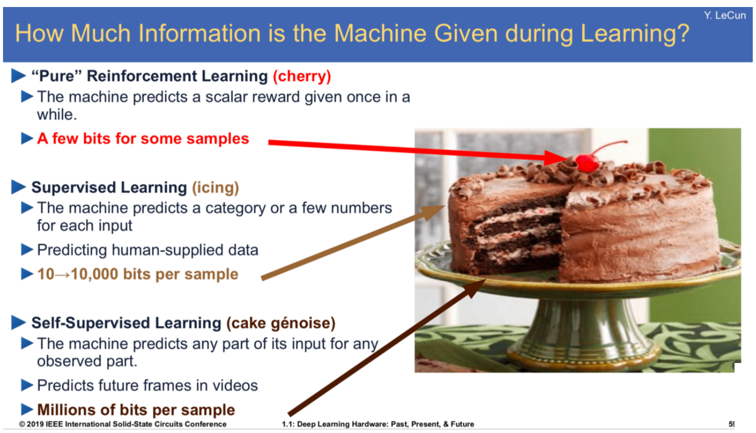
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

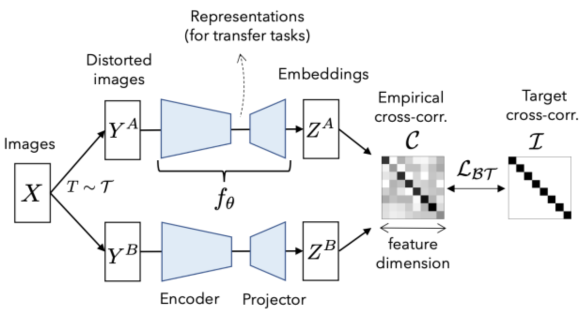
GENERAL STRATEGIES FOR SELF-SUPERVISED REPRESENTATION LEARNING

Theory: 📝📝📝 Practice: 🔨 Complexity: 🚀🚀

Self-supervised representation learning is a set of techniques with the goal of training a backbone network on an irrelevant task where lots of unlabelled data is available. If successful, the model can then be finetuned on different downstream tasks and different domains, requiring less data by several orders of magnitude and, hopefully, obtaining improved performances. You can see why this is a hot topic for both academia and industry. There are two ways of doing it: i) let the network regress hidden or distorted parts of the input or ii) independently distort the same input twice and let the network learn a shared representation that explains both. Either way, these methods have to deal with multi-mode outputs or collapsing representations, and these are exactly the type of issues researchers have recently been busy with.

By working on this topic for your thesis you will have the opportunity to learn more about these fascinating and promising techniques and study some of them to a level of detail required to produce a working implementation. A strong candidate will apply these techniques on real world problems from the company proprietary datasets. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.





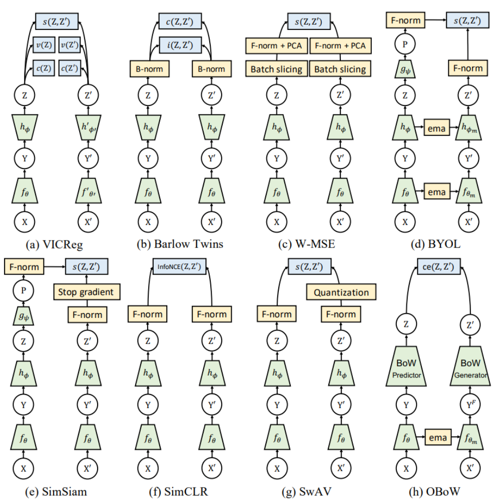


Figure 2: **Conceptual comparison between different self-supervised methods.** The inputs X and

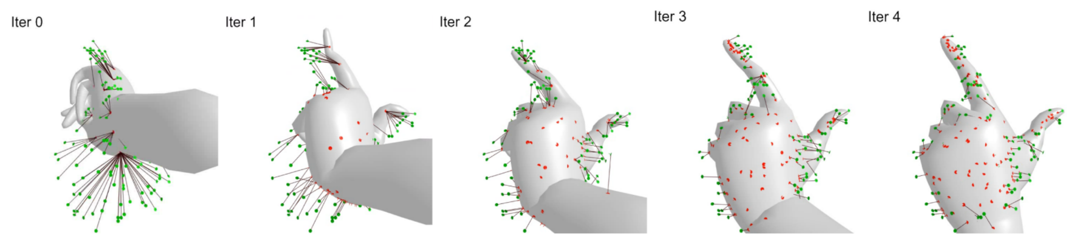
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

DEEP VISION CONSULTING

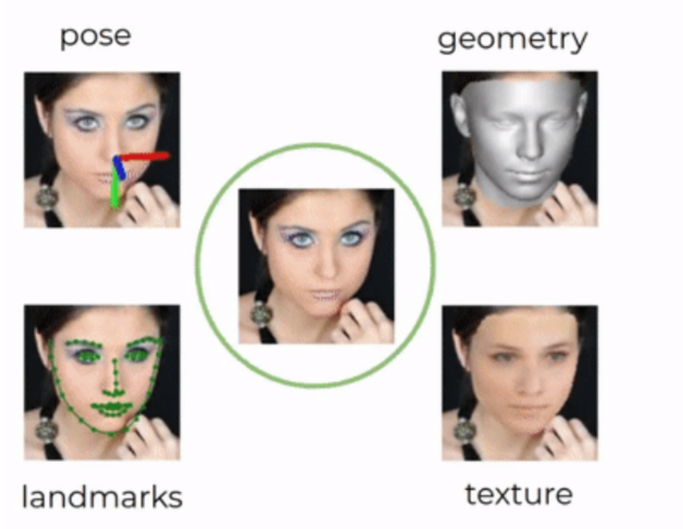
3D MORPHABLE MODELS IN THE ERA OF DEEP LEARNING

Theory: 📝📝   Practice: 🔨🔨🔨   Complexity: 🚀

Fitting parametric models is a top-down approach to understanding the input when you already know what to expect from your data. This is in contrast with classical bottom-up approaches where you start from raw pixels / 3D points and ultimately classify them to discover what you are looking at. Thanks to the model that acts as an additional piece of information, you can often solve problems that would otherwise be too difficult to tackle. Parametric fitting is deeply present in our lives: from Instagram filters that locate the keypoints of your face to Oculus headsets that find and interpret your hands and fingers for control. Deep learning is very appealing to these approaches because many models that were previously too difficult to optimize can now be solved in a single sweep of inference of a properly trained network. Understanding how these methods are able to reliably solve the fitting optimization problem is the real challenge, and it will be the goal of this internship. A strong candidate will also choose and implement the most appropriate approach to solve a real world problem from the company proprietary datasets. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.



(an example of the Phong hand model)



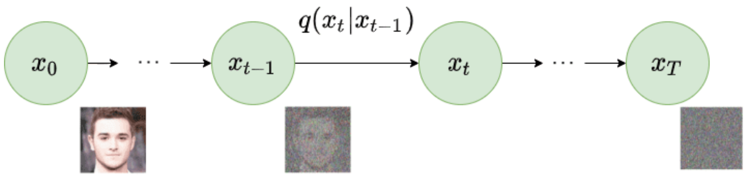
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

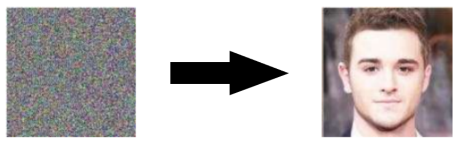
DIFFUSION MODELS: WHAT'S IN IT FOR THE INDUSTRY?

Theory: 📝📝📝 Practice: 🔨🔨🔨 Complexity: 🚀🚀🚀

2022 has been the year of diffusion models in computer vision. Stability AI, Open AI (Dall.E) and Google (ImageGen) are just some of the big players that have either released their pretrained models or given public access to their API. The basic idea behind diffusion models is rather simple. They take an input image and gradually add Gaussian noise to it through a series of steps. Afterward, a neural network is trained to recover the original data by reversing the noising process. By being able to model the reverse process, we can generate new data from noise. This is the so-called reverse diffusion process or, in general, the sampling process of a generative model. Often these models are guided through an additional input sentence in natural language, which biases the reverse process to generate semantically meaningful pictures.

The aim of this thesis is to study deeply these techniques and evaluate their applicability in a few industrial applications that pivot around image generation, like rendering or synthetic training. A strong candidate is expected to train one or more custom models on company proprietary datasets. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.





images generated by Dall.E 2

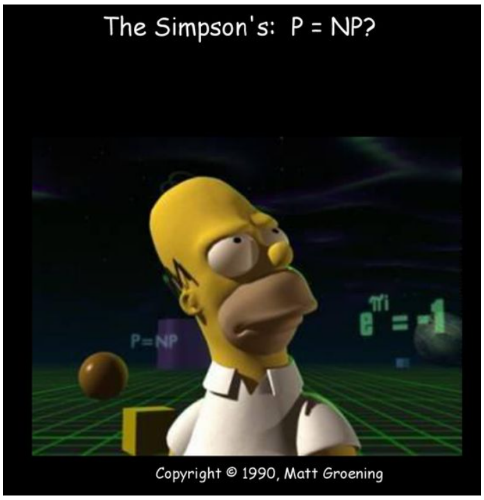
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

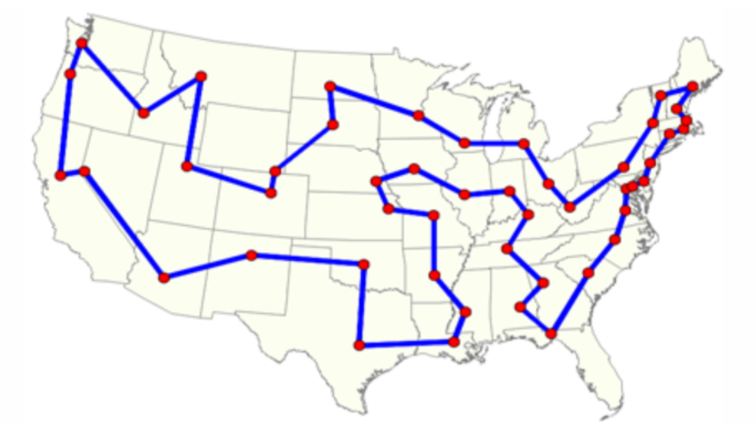
ATTACKING NP-HARD PROBLEMS WITH GRAPH NN

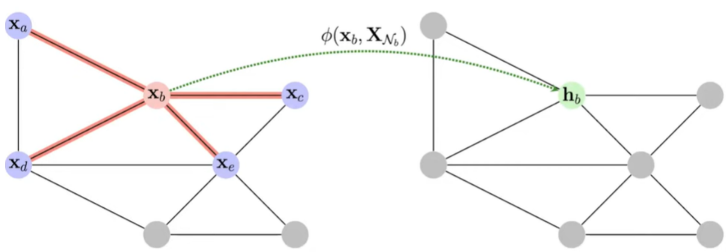
Theory: 📝📝📝 Practice: 🔨🔨 Complexity: 🚀🚀🚀

There exist a number of problem classes for which obtaining the exact solution becomes exponentially expensive with increasing problem size. The quadratic assignment problem (QAP) or the travelling salesman problem (TSP) are just two examples of such NP-hard problems. In practice, approximate algorithms are employed to obtain a suboptimal solution, where one must face a trade-off between computational complexity and solution quality. Since the rise of DL there have been a few research attempts that try to exploit it to approximate such suboptimal solutions.

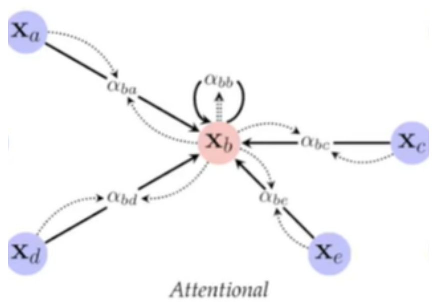
In this thesis, the student will first draw an encompassing view on the state of the art and then propose new graph neural network architectures that learn to solve these classes problem from examples. The student will learn the art and secrets of deep learning on graphs as well as acquiring introductory knowledge in the field of computational complexity. A strong candidate will have the chance to test the proposed architectures on a variety of applications from the company past experiences where NP-hard problems are the core computational bottleneck. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.







$$\mathbf{X}_{\mathcal{N}_b} = \{\{\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_e\}\}$$



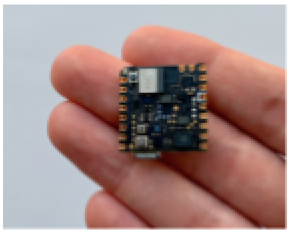
$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j)\psi(\mathbf{x}_j)\right)$$

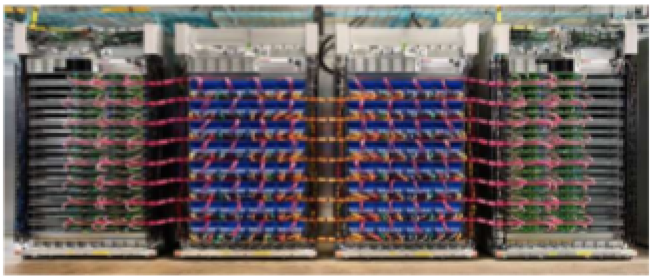
DEEP VISION CONSULTING

DEEP LEARNING INFERENCE ENGINES: THE RIGHT TOOL FOR MODEL DEPLOYMENT

Theory: 📝 Practice: 🔨🔨🔨 Complexity: 🚀🚀

Deep learning isn't all about training and experiments. Every custom trained model must be deployed somehow to production. There are a number of valuable solutions that allow you to run your networks on all sorts of hardware: from tiny embedded microcontrollers to scalable performant cloud machines or even data centers. We refer to these solutions as DL inference engines or runtimes: TFlite, TensorRT, TNN, Onnx, pytorch, tvm, tinyML are just a few examples. But every DL inference engine has its own peculiarities: some run only on the cloud or only on the edge, others try to achieve low latency or high throughput.
Is there a DL inference engine to rule them all?
During this internship you will explore the most modern and promising inference engines to compare them and understand pros and cons of each. A strong candidate will have the opportunity to go from the tiniest microcontrollers up to the biggest cloud providers. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.





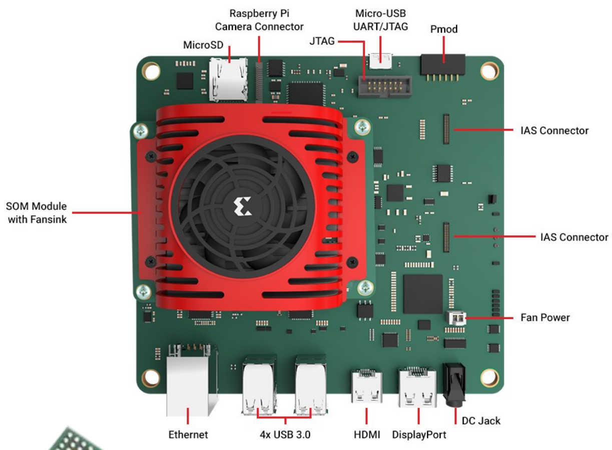
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

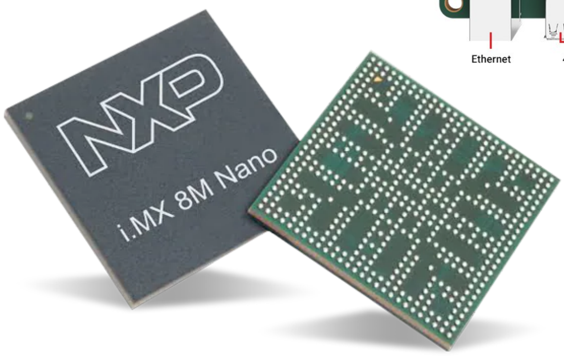
DEEP VISION CONSULTING

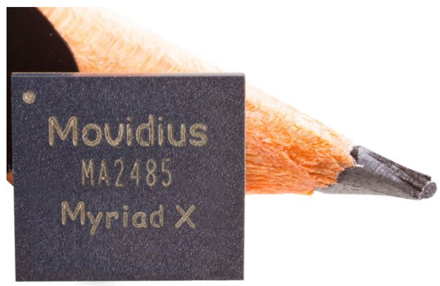
DEEP LEARNING AT THE EDGE: AI ACCELERATORS

Theory: 📝 Practice: 🔨🔨🔨 Complexity: 🚀🚀

Training models is just the first step for a successful DL project. It's the part we learned at school and that we read in the papers - and it's beautiful: unconstrained computational resources, power supply and time. Then, a few things can happen in the future of a model: stay on a PC with a good CPU / GPU, become a cloud service or go out to the real world running on mobile / embedded devices. Very often, and for several different reasons (technical limits, legal limits, etc), data cannot be uploaded to cloud providers and the only viable option is to process the data where it gets produced, i.e. closeby the sensors. However, doing it via regular GPUs is not an option, because of size, cost and reliability reasons. That's why, since the rise of deep learning, semiconductor companies have created embedded hardwares specifically designed to run models at the edge. Nvidia Jetson (TX2, Javier, etc) is the most popular but not the only one: Xilinx UltraScale+ (Kria Som), Intel Myriad, NXP NPUs (iMX8 M Plus) and STMicroelectronics STM32 are just a few remarkable alternatives. The goal of this thesis is to take a few challenging problems from the classical repertoire but solve them with custom architectures deployed on highly efficient and ultra-low power devices. You will learn a lot about how to make a DL model run efficiently on lots of different hardware. A strong candidate will also have the chance to test modern techniques to further improve the inference speed of at the edge models, such as quantization, pruning and distillation. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.







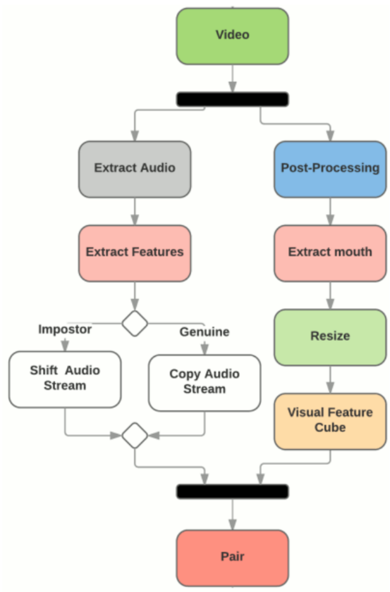
Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

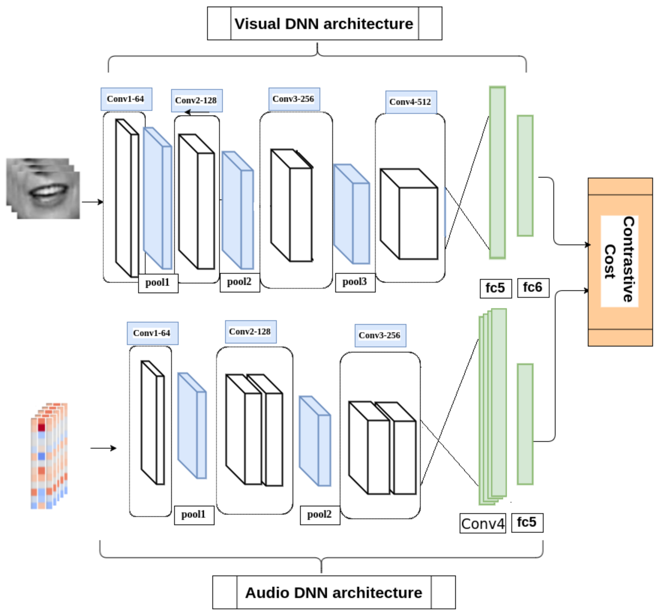
MULTI-MODAL DEEP LEARNING WITH AN APPLICATION TO LIP READING

Theory: 📝📝 Practice: 🔨🔨🔨 Complexity: 🚀

Lip reading from videos is quite a solved problem in computer vision, from a research perspective. Then why would anyone want to do a research internship on this topic? Because you can learn a lot by trying to solve the task from scratch. Look for the right data, build your model by taking inspiration from published work, experiment with scientific rigour and try to compete with the state of the art methods from world renowned research labs. The cherry on the cake for this task is that it comprises data from both the audio and video domain. A strong candidate will also have the chance to exploit self-supervision to leverage captions from YouTube videos and enrich the public training sets. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.

Fig. 5. One-second clips that contain the word '*about*'. Top: male speaker, bottom: female speaker.





BEYOND OPENCV: NVIDIA NPP AND INTEL IPP

Theory: 📝 Practice: 🔨🔨🔨 Complexity: 🚀🚀

OpenCV is a widely used computer vision and machine learning library. It provides thousands of optimized functions to help you solve your computer vision problem. But sometimes it isn't the right tool for your job. It may be too big, or not fast enough, or it's missing that specific algorithm you may be looking for. What if there is another library out there that provides exactly what you need?

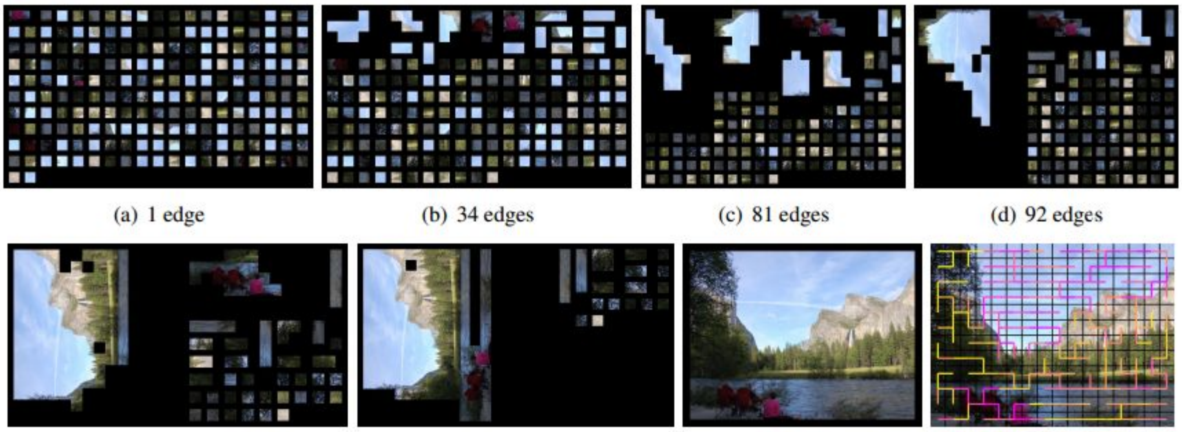
The aim of this thesis is to explore the boundaries of [NPP]() and [IPP](), two high performing vision-related libraries from Nvidia and Intel. You will get to apply these libraries on real problems derived from Deep Vision Consulting experience. A strong candidate will also have the chance to implement new and previously unsupported functions of great use to computer vision for one or both of these libraries. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.

Deep Vision Consulting S.r.l.
www.deepvisionconsulting.com | info@deepvisionconsulting.com | Phone. +39 059 8678417
N. Iscriz. Reg. Imprese di Modena, C.F. e P.Iva 03844340368 | N. REA MO – 422631 | Cap. Soc. €10.000,00 I.V.
Registered Office: Viale Vittorio Veneto N.2, 41124 Modena, Italy
Headquarter: Via Capilupi 21, 41122, Modena, Italy

JIGSAW PUZZLE / RUBIK CUBE SOLVER

Theory: 📝 Practice: 🔨🔨🔨 Complexity: 🚀

Humans have limited attentional bandwidth, approximate visual memory and suboptimal strategies to get to the end of combinatorial problems. That’s why solving a jigsaw or the rubik cube is fun to (some of) us: because we don’t have a naive and efficient solver for these tasks already built in our brains. Conversely, computers shine at things like this. As a computer scientist and engineer you may have wondered what it would take and how hard it would be to develop an automated program to solve famous puzzles for you: if you did, this internship is for you: discover the answer by coding! You will learn how to approach the computer vision part of these tasks, the mathematical abstractions required to formulate the combinatorial optimization problem and, eventually, develop all of this as a mobile application. Strong candidates will also have the opportunity to make their solutions more efficient through the use of machine learning algorithms. The candidate will carry out his research work onsite at Deep Vision Consulting and be tightly supervised by an experienced member of the team.



(a) 1 edge (b) 34 edges (c) 81 edges (d) 92 edges

(e) 156 edges (f) 174 edges (g) Assembled Puzzle (h) Spanning Tree