**Interviewer 1:** [00:00:00] Okay, so I could we start by asking you some general questions about your background. So what is your current position at your workplace?

**Interviewee:** [00:00:11] Yes. I'm a postdoc at [University: Remover for Anonymity]. After PhD spend here as well and I am working on forensics Digital Image forensics team [inaudible]. Yeah, it has something to do with adversarial learning be there but or now involved in deep learning stuff that in this specific field.

**Interviewer 1:** [00:00:42] So how much of overall work experience do you have?

**Interviewee:** [00:00:46] 3 years of PhD, two years of postdoctoral, so five years after my MSc degree.

**Interviewer 1:** [00:00:59] Thank you, and the experience in deep Learning System -how many years of experience you have in that?

**Interviewee:** [00:01:06] Oh, maybe since 2015, so 3-4 years.

**Interviewer 1:** [00:01:17] Thank you. Could you tell us what type of deep learning networks have you developed, implemented, and by that I mean was it supervised and unsupervised, reinforcement learning, things like that?

**Interviewee:** [00:01:27] Yes, I am dealing with supervised, of course, unsupervised as well, GAN, some autoencoders. [inaudible], LSTM, RNN. I developed a lot of architectures but my papers, mainly there are supervised or weakly supervised architectures.

**Interviewer 1:** [00:02:03] Thank you. Which problems were you trying to tackle when using deep Learning Network like image classification speech recognition. You told us something about forensics, but would you give more details?

**Interviewee:** [00:02:14] Yeah, you mean problem in programming or problem in general?

**Interviewer 1:** [00:02:19] Yeah, but what do you want your deep Learning Network to do lets say.

**Interviewee:** [00:02:23] Oh, yeah the first thing is I would have some architecture some framework that works with the more specific framework. For example, Tensorflow framework cannot work with Keras, with the Pytorch. If you have a [inaudible] Theano and you cannot use it if you want to use tensorflow and so this is the main problem - on the net that you have to give a lot of possibility, but you have to change the framework each time. So, this is a problem

**Interviewer 1:** [00:03:12] So what we want to know is that.. Like deep learning networks, they are trying to achieve something. They want to classify images correctly or to be able to recognize speech correctly and you tell us which problems of this kind you're trying to address.

**Interviewee:** [00:03:31] A, Okay. Mainly to classify I spot... I don't know if you deal with forensics or you heard about forensics. Usually we have to classify for example a camera model which has taken a photo or you have to classify if the picture is forged or not, where it is forged, so I have to localise it. So it is a classification, it is like image classification, but the classes are different from the classical object prosecution - the objectives are different, but usually is a classification localization [inaudible] segmentation tasks.

**Interviewer 1:** [00:04:21] Very interesting.

**Interviewee:** [00:04:23] Yeah, I hope so.

**Interviewer 1:** [00:04:25] Could you tell us which programming languages and Frameworks have you been using so far?

**Interviewee:** [00:04:31] No, I use only python sometimes Matlab for old code. And I use Tensorflow, keras, pytorch. Just one time I used Theano. Oh, I saw MXNet net but I have never used it.

**Interviewer 1:** [00:04:55] Okay, so I think our background questions are over. And in this interview we have like one general questions: what are the bugs and problems that you have had during developing these systems. So if you could just start talking about that would be great.

**Interviewer 2:** [00:05:15] Even small technical things, so we're interested in everything you can come up with.

**Interviewee:** [00:05:28] The biggest problem when we deal with Deep learning frameworks is the compatibility of version, you know deep learning frameworks are very young and are very fast-growing. You will find a new version, maybe two new versions each six months. And sometimes they have to they have to change a lot of interfaces and the back compatibility is not always a guarantee. And so the biggest problem that I find is to use all the code with the new version of the framework.

For example, I have a lot of error because I have a new version of the framework and I have to install the older one, but it is too old, you have to find it. So the biggest problem in deep learning for me is this. Other kinds of problem are more practical problems. For example, how to manage the memory and all this stuff, but it is a problem of the platform. For example, if you have a small GPU you cannot fit the [inaudible]

**Interviewer 1:** [00:07:16] So when you have out of memory problems, switching to a better machine in the only solution you use or how [inaudible].

**Interviewee:** [00:07:25] Yeah. No, of course better machine, a better GPU, the number of GPUs is one way to solve the problem or you have to code, you have to reduce the number of images. So there are solutions, sometimes it doesn't work with a minimum of setting.

**Interviewer 1:** [00:07:54] Okay, anything else. about problems and bugs?

**Interviewee:** [00:08:16] Sometimes there are hidden bugs, because sometime the thing doesn't work but you don't know why, so there are not bugs, that are things that just doesn't work and it is a big problem - you cannot debug some framework. For example, tensorflow, It's very powerful, but you cannot debug it you cannot go line-by-line in the code to see what exactly. So you just have an error or you have some strange results and you cannot understand where to put an effort to change the code.

**Interviewer 1:** [00:08:57] Can you give us an example of an error of the or strange results that you got from tensorflow?

**Interviewee:** [00:09:03] Yeah, for example, sometimes you have problem of dimensionality. We cannot fit this vector in the other but you don't know in that point which is the size of your vector because you cannot go through the code. For debugging it is very very difficult. So usually you have this kind of problem or you cannot sometime.. For example, in tensorflow mainly.. So I moved by pytorch. I have to think about that. Because if you [inaudible] Deep learning, you know this kind of problem.. Sometimes they change the way to pass the parameter.

**Interviewer 1:** [00:10:31] Okay

**Interviewee:** [00:10:32] And so the error is there. But you don't know why because it is just a version problem, this match, and you see only the results that are broken and you will know that there is an error. So with deep learning sometimes the error is not a tracked. If you mean about error code, the recovery, the main error code compatibility with older version and another problem is that, for example, recently I had a problem with json version. I had old code with the [inaudible] and they cannot be read by the code because I had to use an older version so it is it just this kind of problem I meet with the code.

**Interviewer 1:** [00:11:44] So when you are developing your deep learning systems, did you collect your own data or do you use some existing training in data?

**Interviewee:** [00:11:53] Both. With deep learning, if you deal with computer vision stuff, usually you have a lot of [inaudible] sets already done but in forensics there aren't so many that sets and sometimes we have to collect ot create synthetic datasets.

**Interviewer 1:** [00:12:19] So, how did you collect your data or how did you create the syntetic..

**Interviewee:** [00:12:26] It depends. For example, recently we had a project with [Academy: Removed for Anonymity] here and we borrowed some devices and taken some photos, we went to the street to take some photos to collect the data, or we downloaded from internet. And since we try to understand [inaudible] images are manipulated or not, so we manipulated a bit manually with Photoshop, for example, I have a paper with GAN detection. So we detected if images are GAN generated or not. And so we collected a lot of dataset trying to replicate the results of other papers. That's the way we collected data.

**Interviewer 1:** [00:13:28] So in your data set the only labeling that you have to do is whether the image has been manipulated or has not been manipulated?

**Interviewee:** [00:13:38] It depends, usually the labeling is at pixel level so you can do it manually or automatically or is it manipulated or not or is a [N class/ Mplus] problem. So we have M model of cameras, we have to understand if a photo come from one on these cameras. Or sometimes, even the blind classification so with clustering so we have no label that is it, we [inaudible] it.

**Interviewer 1:** [00:14:23] And when you have to provide manual labels, do you do the labeling yourself or is there anyone else involved?

**Interviewee:** [00:14:34] Oh, usually by myself, or the people who work with you.

**Interviewer 1:** [00:14:41] Do you usually check if you have an agreement in your labeling?

**Interviewee:** [00:14:48] No, because usually I know the exact label.

**Interviewer 1:** [00:14:54] Okay

**Interviewee:** [00:14:56] Not personally, in my group, recently, we had even a human agreement of just to understand if a human can beat a deep learning. It's not a label in this just to check if..

**Interviewer 1:** [00:15:18] And what were the results for that?

**Interviewee:** [00:15:22] Sometime you can be fooled [inaudible], for example a GAN generated images are difficult to be understood for human because they are very very photorealistic.

**Interviewer 1:** [00:15:40] So do you usually pre-processe the connected data in some way or do you just use it as it is?

**Interviewee:** [00:15:54] Yes. Yes, sometimes use a some noising algorithm but we use the noise part of the... Usually [inaudible] to use the clear part and use the noise, but the other kind of pre-proposed are standard ones: compression or resizing, just normalization to have the data in the same range.

**Interviewer 1:** [00:16:30] Tell us about the case where you did not do some kind of pre-processing and try to train, you got into trouble. You went back. You applied some pre-processing.

**Interviewee:** [00:16:45] Yeah, exactly, student did not propose correctly the image. UIn that case it was a just a [inaudible] to understand. When you open an image., Usually you have a 022 that 255 range. Okay, by the Deep learning [inaudible] a smaller array from 0 to 1 or minus 1 to 1 to be very close to 0 and [inaudible] the problem with the gradient back propagation. And in that case for example without any proposal [inaudible] if you forget to normalize the image, you know a very strong result.

**Interviewer 1:** [00:17:39] So could you remember any other kind of problems related to training data that you ever had?

**Interviewee:** [00:17:46] Yeah, usually the problem is in setting some hyperparameter - learning rate or the number of images in a batch, number of epochs. So sometimes they are too much and you go in the overfitting and you do not know and because of the lazy programming you don't check and you just run an experiment. Nobody is looking for it so the mean main problems are these.

**Interviewer 1:** [00:18:40] So maybe you could give us more details, for example, with learning rate what kind of problem you had?

**Interviewee:** [00:18:52] The standard optimiser is stochastic gradient descent [inaudible] and is very very sensible to the learning way. Okay, and you have to set it very carefully. But if you use a Adam, for example, that is another Optimizer that uses the [secondary?] rate and it just first is a more stable.

So you have to think about the learning rate, but usually it helps you too resolve a lot of problem. For example, there was a code for a camera model identification of some people. I try to replicate the results and nothing worked. It was just a random guess I just changed the SDG to the ADAM, for example, or other grad in that case and everything works correctly just because sometimes you have to set when [inaudible] very for example change a lot if you set the zero one or zero zero one, for example, and it changed a lot. You have to be careful to have a smooth learning rate of Decay. But for all these problems you do not have the error message.

**Interviewer 1:** [00:20:58] Then, you mentioned batch size, that you should change that one.

**Interviewee:** [00:21:08] Usually, batch size dimension problem is the more is the best usually battle.

Sometimes you have a big architecture and you just want GPU. Or you don't want to write a lot of code to use two gpus in just one time. You try to hit it, you try to increase it to fit your architecture in [inaudible] your GPU. I never experienced too low or too small batch. It was too tiny.

**Interviewer 1:** [00:22:02] Ok, so you usually increase by size rather than decreasing it.

**Interviewee:** [00:22:08] Usually you use a 32 or 64 and go ahead but, for example, the my last paper. I deal with the whole image classification, so without resizing and I had to tile images in more batches and I couldn't fit one big image, 5,000 x 4,000 in one batch.

So I had to split it in several beaches and use lot of code to keep the back propagation right. Sometimes you have to deal with these problems because what I experienced that the direction of this framework are to solve this kind of problem, for example, now, there are checkpointing in some stuff for helping the developer to write as less code as possible for this kind of problem, for [inaudible] memory, to [inaudible] gradient space, to fit the memory.And Steve's to fit your memory and so on. The problem of all deep learning tools are that they are too young and they are not [inaudible] because you have a bigger problem and they are not ready for you.

**Interviewer 1:** [00:23:55] And you mentioned also the number of epochs overfitting, do you do that? [inaudible, most probably: could you give us more details on that]

**Interviewee:** [00:24:05] Sorry?

**Interviewer 1:** [00:24:05] You mentioned the number of epochs..

**Interviewee:** [00:24:08] Oh. Usually we use a validation set to set the number of epochs. There are some problems that are harder where do you need thousands of number of epochs and others that just [inaudible], for example, it dependds.

**Interviewer 1:** [00:24:40] You also said that you use existing data sets, checking data set. So could you mention which ones you have used?

**Interviewee:** [00:24:48] Ok,ok..

**Interviewer 1:** [00:24:48] If you remember, of course.

**Interviewee:** [00:24:54] I used, for example, the ImageNet, MNIST, CIFAR [inaudible] as computer vision one, but I used the Dresden data set that is the dataset of forensics. Another is called Lesion, for example Socrates, I downloaded GAN generated images so they are not officially a dataset but we created it by collecting small data sets or I used the dataset for Iris Biometrics fingerprints. There are lots so..I have a list, if you want, I can send you a email..

**Interviewer 1:** [00:25:45] So did you ever have problems with existing datasets, training using them?

**Interviewee:** [00:25:56] No, no. No, I downloaded and that's it. They are very well organized. When we create a new data set so we put a lot of effort because it's much more difficult.

**Interviewer 1:** [00:26:22] Okay, so just to summarize this part of our conversation any problems or bugs related to training data as collected by you or the ones that you have downloaded that come to your mind right now.

Just to summarize...

No,no, if you have anything to add about this.

**Interviewee:** [00:26:47] To summarize, the problem of data is always very strict in deep learning. Usually the problem is we have not enough data. In forensics, for example, so we have to use a lot of effort for generating new dataset. For example, we used a script to download images from Flickr or Facebook automatically to collect as much data as possible for a challenging camera mode alleged issue because we had just names of cameras, but very few images and we have to reach 1 million of images for training our algorithm and we had a lot of effort it it. Sometimes the problem is [inaudible] typical problems, for example. object classification, missing data segmentation. There are a lot of datasets, they are well organized, they fit in your problem. The problem is to use a different solution when you have the problem of a collecting data is very big problem, it is up to you.

**Interviewer 1:** [00:28:35] Thank you. So let's now talk about the structure of the models you have been using like number of layers, type of layers. Did you ever had problems related to that?

**Interviewee:** [00:28:46] So I use from MLP to 1000 of layers..No, I haven't [inaudible]..

**Interviewer 1:** [00:29:02] So the model that you usually use has a lot of layers?

**Interviewee:** [00:29:06] [inaudible] yes. For example, [inaudible] I use a [inaudible], DenseNet [inaudible].

**Interviewer 1:** [00:29:14] Did you ever have to change anything about this model structures or you just always use them as they are.

**Interviewee:** [00:29:25] Yes, sometimes we change it.

**Interviewer 1:** [00:29:28] Can you tell us about some changes you made and why?

**Interviewee:** [00:29:32] Yeah, sometimes we try to change the type of padding and the stride because the in forensics you use a stride so down-sampling the image, [inaudible] is not a good idea because you use that information use a micro button. So if you are down-sampling something [inaudible] something or throw away some pixels, maybe that pixels are important we use the [t I frequency??] Part of the frequency domain. So if you down-sample and you throw away [inaudible]. So we found that in our problem this kind of approach is better, but usually when you have a lot of data and because problem is to use the architecture that are already trained for training [inaudible] different task. Usually we try to use network that is trained and we have to just fine tune it.

**Interviewer 1:** [00:31:02] So fine tuning includes only changing the hyperparameters or anything else?

**Interviewee:** [00:31:08] Usually hyperparameters or, of course, the loss, thre kind of a classifier that is different. Because you know we see that.... We even used the handcrafted architecture for some kind of problem. When you I use already given Network, for example, Xception network, it is very powerful recently, we change some hyperparameters, the kind of classifier the kind of loss.

**Interviewer 1:** [00:31:56] What can you tell us about the change in terms of classifiers, [inaudible] from one to another and [inaudible].

**Interviewee:** [00:32:06] For example, we use a different loss. For example, if I see [inaudible] we are doing this. If we have a classifier with different kind of class we have to you have to change a layer and it is natural for fine tuning, but sometimes the loss, the cross entropy loss is not enough because you need [?] even reconstruction loss. In specific cases you need an additional loss. Or that the classifier is not an MLP, but you just throw away the end MLP, you put an SVM so it's kind of changing the network.

[mixed, inaudible]

**Interviewer 1:** [00:33:04] So change the predefined loss functions, but from one to another, right? That find [inaudible]

**Interviewee:** [00:33:12] Yeah sometimes even a crafted. Slightly different from the....

**Interviewer 1:** [00:33:20] So you use the custom written loss functions, right?

**Interviewee:** [00:33:25] Yeah

**Interviewer 1:** [00:33:25] Could you tell us about any problems you have faced using predefined or custom written loss function.

**Interviewee:** [00:33:32] The predefined works.

**Interviewer 1:** [00:33:35] So with the predefined you only have to find which one is..

**Interviewee:** [00:33:41] Yes.

**Interviewer 1:** [00:33:41] Yes,with custom wrtitten?

**Interviewee:** [00:33:44] Sometimes you [inaudible] error, right? Yeah, the problem is in is not called error. So they work but you don't know why [inaudible] not. They don't give you the result, for example, we used one particular loss or diffuser. We change it from distant to correlation.

**Interviewer 1:** [00:34:16] [inaudible] What?

**Interviewee:** [00:34:19] You know, if you use a [??] Network

**Interviewer 1:** [00:34:22] yeah

**Interviewee:** [00:34:23] you start between two batches [inaudible]. In that case we need correlation [inaudible]. It is not so different. So you are looking at the direction, if you know at my experience just looking the angle but not the distance. Okay, in that case, it was a little bit difficult .. It wasn't difficult. But, you know, we need some effort to set up everything but it works. It iw a bit difficult in the beginning because you have to undestand where to change the code. Usually we reused the code and changed it because our approach is to use water was already done and the to change it.

**Interviewer 1:** [00:35:29] So from what you're saying I get the pictures that you have big models. So your training time should be big, that's the impression that it gives it. Is that true?

**Interviewee:** [00:35:43] Sorry?

**Interviewer 1:** [00:35:45] So you were talking about that you're handling big pictures when training different..

**Interviewee:** [00:35:49] Yes, Yes..

**Interviewer 1:** [00:35:50] I get an impression that your training time should be long?

**Interviewee:** [00:35:54] Yes.

**Interviewer 1:** [00:35:55] Yes. So, where do you train your models?

**Interviewee:** [00:36:00] How many time?

**Interviewer 1:** [00:36:04] Like on which kind of machine, do you use specific hardware?

**Interviewee:** [00:36:08] A, ok. Yes. Now we have two server with the 2 Tesla [inaudible] each, so we have four GPUs in two servers and we have other servers with smaller GPUs. But usually I use one GPU a time because we are a group and everybody has to use GPUs. It took even several days to train [inaudible] big images.

**Interviewer 1:** [00:36:46] Can you tell us about an example of an improvement that you have done in your model which reduced your training time in a significant way.

**Interviewee:** [00:36:59] Usually in case of my master people. The biggest problem was the time I took to read image from the [file?] because all the dataset cannot fit in RAM and so you have to read it from the files each time and I'm and yeah, the biggest.... And the framework has the loader that uses threading, [inaudible] several workers and it is [inaudible], it gives a strange error. Just without an error message. My computer just stop, just doesn't work anymore. I have to reboot it. Without any error message without nothing. So I had to write code by myself to read image in parallel way without having a problem and that was the biggest [inaudible] I added in that case because okay, when you have a big architecture and you give a lot of data you expect that it takes time. But if the time is the one for reading images you waste a lot of GPU time. Because the GPU is waiting for you, for data.

**Interviewer 1:** [00:39:00] So could you tell us about some problems that related to Hardware that you have ever encountered?

**Interviewee:** [00:39:07] Some problems? sometimes your voice is..

**Interviewer 1:** [00:39:11] I will repeat. Did you ever had problems related to Hardware like errors, crashes, because..

**Interviewee:** [00:39:29] Just this kind of problem that I think is [inaudible] related because I don't know why, but the server doesn't respond because the virtual memory was full. You know, you expect that it killed your process and go away and go ahead. But in that case my processor wasn't killed and but I cannot access to my server. The only way was to go to my server rack and reboot my machine. It was the only other related other sometimes, yes, when you install new GPU you can have a problem with the driver, but it is software problems, not hardware problem.

**Interviewer 2:** [00:40:28] And you told us that you mainly use one GPU. But when you use multiple ones, do you ever encounter any problems related to concurrent execution of your code?

**Interviewee:** [00:40:42] Usually when we use multiple deep you we have to write our code by ourself because we haven't [inaudible] the link and some virtualisation. And it is very very boring and we we had a problem because in that case, we usually we try to avoid this kind of situation because it is it's really boring and usually I try to split into 2 part the model, the architecture in the in a way that you haven't this kind of program. They have a just one single point of Union and they are [inaudible] two different tasks and we use the results. Usually we try to do something for avoiding this kind of problems because they are very very very annoying and we haven't time to do it.

**Interviewer 1:** [00:42:03] So is there any other kind of problems that we haven't mentioned today? It come to your mind that you could ask.

**Interviewee:** [00:42:12] That's probably my see that there is a form, maybe I can think about it and the kind of problems related to deep learning?

**Interviewer 1:** [00:42:25] Yes. Yes, exactly. Do you remember any nasty bugs that took a lot of your time or any error message that you think you get too often?

**Interviewer 2:** [00:42:34] Yes, even the small ones, but you just get them from time to time.

**Interviewee:** [00:42:43] No, just warning messages sometimes happen and the annoying part is when a lot of message that say that remind you that the cold you are using will be deprecated after a version that will come next year and you always know that you should change something in your code but you don't do it. But yeah, usually these kind of problems. And what other kind of error..no, maybe in the beginning but it was my bad coding, you know when beginning and you haven't experience, but I don't remember. Yeah, of course problem of dimensionality, problem [inaudible] as I mentioned.

**Interviewer 1:** [00:43:51] Dimensionality, is it about the dimension of the input [inaudible]?

**Interviewee:** [00:43:55] Yes, in the past. Recently the networks have a global average pooling at last so you don't know what is they are doing but it just reduces your image to 1 to 1 vector, 1 to 1 to n vector. In the past when the architectures were more handcrafted, if the architecture worked with for example, 128 by 128 image, you can change it because, you know, the dimensionality of the feature vector was a sum of convolution stride and at last you had to [inaudible] that expect 4000 long length vector and it was better if you change something in the input you expect to change something in the output but it was an old problem, now i do not face this kind of problems

**Interviewer 1:** [00:45:10] okay, so if you have anything else to add you're welcome.

Ok, if during these days

**Interviewee:** [00:45:17] I have some something to say you I will write you a mail. Remembering everything that during these five years happned with me is very difficult.

**Interviewer 1:** [00:45:31] Yes, of course. We know that

**Interviewee:** [00:45:33] Yeah, that's why I said about the form because maybe you have more time. I could take half an hour to remember..

**Interviewer 1:** [00:45:46] Yes, if anything comes to your mind and you could send us an email, that would be very useful. Thanks a lot for doing this, for your time.

**Interviewer 2:** [00:45:56] Thank you very much

**Interviewee:** [00:45:58] Say hi to Vincenzo!

**Interviewer 1:** [00:46:02] We will, yes.

I hope to her about your project and your results soon.

We will definetely the inform you about that.