



# 琶洲实验室 | 黄埔

## PAZHOU LAB · HUANGPU

# 实习报告

课题名称：基于 Stable Diffusion 及 YOLOv5 的工业图像缺陷检测

实习地点：琶洲实验室（黄埔）

学生姓名：邬成浩

指导教师：徐晨博士

2023 年 9 月 14 日

## 目录

一、课题分析.....	3
(一) 课题背景.....	3
(二) 课题目标.....	3
(三) 相关文献.....	3
二、整体思路.....	4
(一) 理论基础.....	4
(二) 解决路径.....	6
三、实际方案.....	6
(一) 准备环境及数据集.....	6
(二) Stable Diffusion 扩充数据集.....	7
(三) YOLOv5 模型训练.....	7
(四) 模型推理.....	7
(五) 模型评估.....	8
四、问题及应对方法.....	8
(一) 环境与文件相关问题及应对.....	8
(二) Stable Diffusion 相关问题及应对.....	8
(三) 模型训练相关问题及应对.....	9
五、结论.....	9
(一) 结果.....	9
(二) 未来改进.....	10
六、感想与总结.....	10
(一) 领略了请教与科研手段的重要性.....	10
(二) 高端且实用的知识.....	11
(三) 知识的实际运用及经验.....	11
(四) 团队协作共同处理解决难题的重要性.....	11
参考文献.....	12

## 一、课题分析

### （一）课题背景

如今，工业发展愈发昌盛，对于工业产品的质量要求也日渐提高，因此，检测缺陷样本成为了工业生产中不可缺少的一环。传统的缺陷检测依赖人工，原因是人具备统合经验常识以发现缺陷样本的能力，即使是从未见过的缺陷样本，人的视觉系统也拥有感知到异常的固有能力，也就是举一反三的泛化能力。但是，传统的人工检查往往依赖主观判断与工作经验、耗时耗力以及容易出现差错。重复、单调但严格的检测任务更适合自动的检测系统。因此，引入自动化的图像缺陷检测是十分必要的，对于更好的检测系统的需求也在提高。

然而，工业缺陷样本存在缺陷多样、缺陷微妙及小样本的难点。首先，工业缺陷样本是多种多样的，各种缺陷的形状、颜色、大小和位置均有很大不同，这为检测异常添加了重重阻力。其次，因为工业生产逐渐高质高效的特点，工业缺陷样本的异常大多都是较难识别且微妙的，比如擦花和较浅的损伤。最后，工业缺陷样本具备小样本的特点。因为如今工业生产的标准化、规范化、条理化及简单化，缺陷样本占总样本的比例较小。这导致，即使深度学习能够依赖大量标注数据将模型训练出较好的效果，收集大量的工业缺陷样本是非常困难的。而且，工业缺陷样本的类别分布并不平衡，一些缺陷类别的样本量相对较少。

此次实习旨在利用 Stable Diffusion 技术结合 YOLOv5 的目标检测框架，通过扩散模型（diffusion model）来生成更多的各类缺陷图，填充小样本使得工业缺陷样本成为具备一定数量的数据集，从而通过 YOLOv5 训练并改善模型进而完成工业缺陷目标检测。

### （二）课题目标

本次实习的目标是学习并掌握深度学习相关知识，设计且实现一个能够在工业产品的图像中，准确检测出缺陷样本，标注缺陷相关信息（如位置和类型）的工业缺陷检测系统。

### （三）相关文献

在深度学习领域，工业缺陷检测是工业生产关键的一环。不但可以提升产品质量，而且可以加快生产效率和减少生产成本。随着深度学习领域的兴起，各类研究层出不穷，基于深度学习的目标检测方法在工业图像缺陷检测方面也因为优异的性能而得到了广泛的运用。目前比较经典的目标检测方法包括单阶段和二阶段目标检测算法。其中，单阶段目标检测算法主要包括 SSD (Single Shot MultiBox Detector)<sup>[1]</sup>、YOLO 系列(You Only Look Once)<sup>[2]</sup>和 RetinaNet<sup>[3]</sup>等。单阶段目标检测算法的优势在于速度快，只需要一次提取特征。在 YOLO 系列方面，Zhang 等提出了一种增加-means++算法和 Depthwise Coordinate Attention (DWCA)机制

以改进 YOLOv5 算法从而提升检测头盔佩戴的新算法<sup>[4]</sup>。Wan 等则利用多层特征金字塔 (multiple layers of the feature pyramid)，多探测头策略 (multi-detection-head strategy)，和混合注意力模块 (hybrid attention module) 提出 YOLO- HR 从而提升光学遥感图像目标检测的效果<sup>[5]</sup>。然而二阶段目标检测算法会在输出前的中间步骤里选出候选框，虽然速度变慢但是提高了精度。二阶段目标检测算法主要为 Faster R-CNN<sup>[6]</sup>，Mask R-CNN<sup>[7]</sup>等。Zhong 等提出了基于深度学习的自动检测多种疾病的架构—AC- Faster R-CNN<sup>[8]</sup>。这种模型通过扩展卷积和变形卷积的结合，更好地捕捉病变的多尺度信息。对于文献的研究发现，前者集中于将模型进行改进和优化，但由于现有数据集中某些种类的缺陷样本偏少，改良模型的方法所能达到的效果有所局限。因此本次实习集中于通过扩散模型填充数据集以改善数据集分布，从而提升模型的效果进而改善工业检测的准确率。

在图像生成领域，除了传统的图像增强方法外，还有生成对抗网络 (GAN)<sup>[9]</sup>和较新的潜在扩散模型 (latent diffusion)<sup>[10]</sup>。Hui Zhang 等提出异常增广的方法，将缺陷图上的缺陷与对比的正常图相结合从而生成新的缺陷图<sup>[11]</sup>。Niu 等提出使用生成对抗网络的图像绘制，生成的缺陷区域通过缺陷蒙版进行控制，从而控制缺陷强度和缺陷区域<sup>[12]</sup>。扩散模型不单单在图像生成领域有亮眼的表现，在异常检测领域也有竞争力。Victor Livernoche 等发现扩散时间概率模型 (DTPM) 推理时间和结果都优于去噪扩散概率模型 (DDPM)，也在实验中证明了基于扩散模型的异常检测具备竞争力<sup>[13]</sup>。Julia Wolleb 提出了一种基于去噪扩散隐式模型的弱监督异常检测方法，降低了训练的复杂性并尽量保留图像中的细节以应用于医学方面<sup>[14]</sup>。通过对文献的研究，可以发现前者对 GAN 图像生成的研究较为全面，扩散模型助力目标检测的模型提升也有先例。因此本次实习的目标是使用扩散模型进行图像生成，以改善数据集，而后使用深度学习的方法进行模型训练。

## 二、整体思路

### (一) 理论基础

Stable Diffusion 是基于潜在扩散模型 (latent diffusion) 进行图像生成。基于论文 High-Resolution Image Synthesis with Latent Diffusion Models<sup>[10]</sup>，基本框架如下：

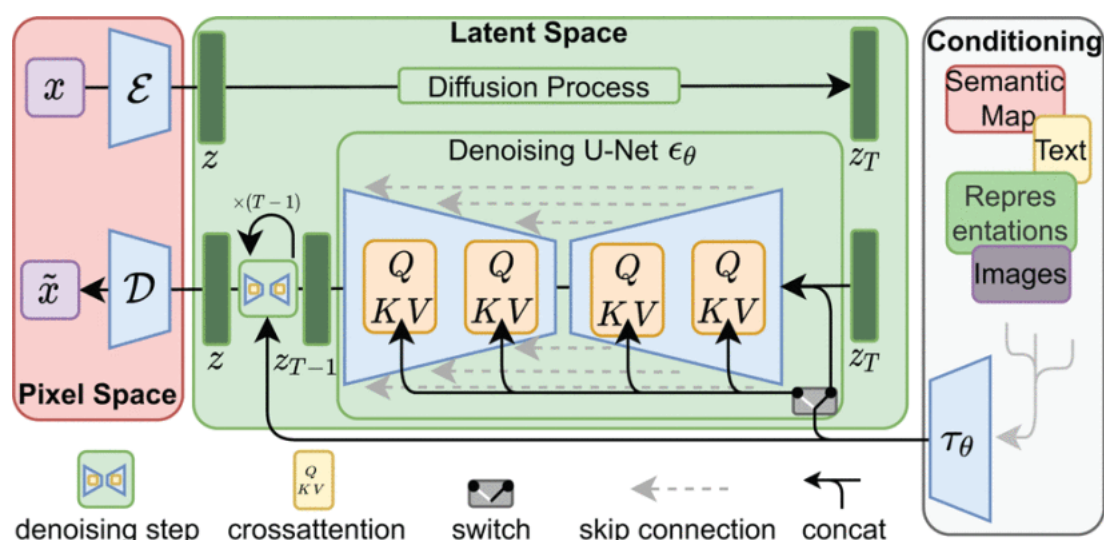


Figure 1 Stable Diffusion Structure[10]

首先是感知压缩(Perceptual Compression)的过程。在像素空间(pixel space), 首先会训练好一个自编码模型 (AutoEncoder), 包含一个编码器和一个解码器。如此便可通过编码器把图片压缩到潜在空间(latent space), 然后最后在让解码器把图片出到像素空间中。通过这样的压缩方法, 扩散模型可以生成细节更多的图像。

然后在潜在空间进行扩散过程(diffusion process)和迭代去噪操作(denoising), 并利用交叉注意力机制(cross-attention)实现多模态训练, 令条件输入成为可能, 譬如输入文字或图片来控制图片的生成。其中, CLIP text encoder 提取输入的文字并通过交叉注意力的形式送入 UNet 中。UNet 则是扩散模型的主体, 用来实现引导下的潜在生成。潜在空间的好处是能够减少计算的复杂度, 使得这个方法可以应用在消费级别的硬件资源上面。这样一来既解决了扩散模型反复迭代导致的计算量的缺点, 也保留了扩散模型相比其他方法(如 GAN<sup>[9]</sup>)在生成质量方面的优势。

YOLO 首次提出时, 基本思想是将物体检测作为回归问题进行求解, 在一个单独的 end-to-end 网络上, 进行原图的输入和目标位置大小的输出。基于 YOLO 的论文<sup>[2]</sup>, 其网络结构如下:

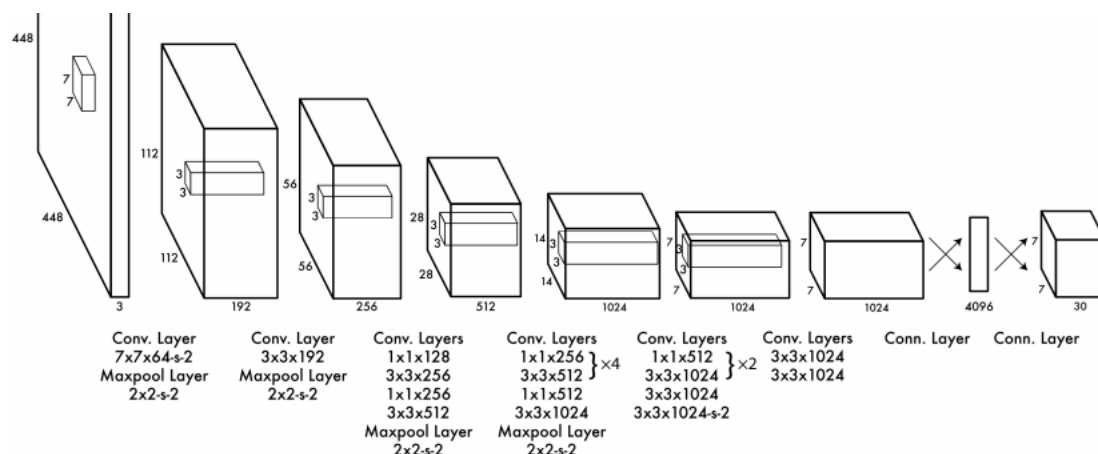


Figure 2 YOLO Structure[2]

检测网络包括了 24 个卷积层和 2 个全连接层，卷积层提取图像特征，全连接层预测位置与给出类别概率。然而 YOLOv5 产生了些变化，Focus 模块换成了 6\*6 的普通卷积层，效率更高。而且将 SPP 优化为 SPPF，也就是空间金字塔池化，将并行改为串行，保证输出相同的同时加快了许多速度。

YOLO 从骨干网络（backbone）开始提取特征，颈部网络（neck）融合浅层图形特征和深层语义特征，最后有头部网络（head）完成检测的模块操作。图形特征指单纯体现在图形上的特征，如颜色和轮廓。语义特征则为更深维度的特征，比如天空是哪个部分等等。最后检测时，YOLO 预测置信度并进行框回归，以完成检测任务。

## （二）解决路径

基于工业缺陷样本存在小样本的特点，提出使用深度学习的数据增强来改善数据集的数量与类别平衡，以提高模型的效果。因此提出假设，通过 Stable Diffusion 生成缺陷图增强数据集可以提高模型的效果。为了验证该假设，首先需要用原数据集对模型进行训练，而后进行模型推理以得到原模型的模型效果。而后，通过 Stable Diffusion 生成缺陷图增强数据集，再进行模型训练得到一个新的模型，然后将新模型进行推理得到新模型的效果。最后对比新旧模型的效果。如果新模型的效果较好，则该假设可以得到验证。

## 三、实际方案

### （一）准备环境及数据集

首先，通过 Anaconda-Navigator 下载 pycharm，然后创建账号并登陆，创建一个项目，激活 python 的虚拟环境。

然后，整理数据集，分出训练集、验证集和测试集。如果是多分类模型的数据集，需对每个类别的数量进行统计，确认哪些类别相对较少。将整理好的数据

集放在项目中。

## (二) Stable Diffusion 扩充数据集

在 GitHub 下载 Stable Diffusion webui 并将其启动，选择 img2img，选择想要进行填充的缺陷类别并从中选中缺陷图作为原图，调整参数如重绘幅度（Denoising Strength），出图量与采样方法（Sampling method），而后进行生成。

生成数量取决于该数据集的分布，尽量使各个缺陷类型的数量平衡。在挑选可用的生成图时，尽量避免与原图过于一致以及与现实缺陷过于不符。如果太多相同的缺陷图存在于数据集中，模型的泛化性和鲁棒性会下降，模型可能会难以准确预测新的数据。如果生成的缺陷图是现实中较少出现的，该生成图不具备实际意义，对模型亦无好处。以下是多分类的原有数量及生成数量表格。

焊盘Abnormal数据	可移动异物	异物	异色	损伤	氧化	漏铜	焊盘凹凸	脏污	全部类别		
原有数量 (张)	38	62	111	567	357	80	78	506	1799		
生成数量 (张)	268	240	200	0	0	220	230	0	1158		
总数量 (张)	306	302	311	567	357	300	308	506	2957		
生成占总数比例 (%)	87.58%	79.47%	64.31%	0%	0%	73.33%	74.68%	0%	39.16%		
基材Abnormal数据	假性漏铜	基材划痕	基材小点	基材损坏	异物	擦花	气泡	漏铜	脏污	边缘损坏	全部类别
原有数量 (张)	1139	303	248	67	154	333	169	450	88	145	3096
生成数量 (张)	0	0	60	240	150	0	140	0	220	150	960
总数量 (张)	1139	303	308	307	304	333	309	450	308	295	4056
生成占总数比例 (%)	0.00%	0.00%	19.48%	78%	49%	0.00%	45.31%	0%	71.43%	50.85%	23.67%

Figure 3 Numbers of abnormal data

本次实习扩充了二分类模型的数据集和多分类模型的数据集。二分类数据集是对比挑图后的结果，共有近两万张。多分类数据集有焊盘数据与基材数据两大类，其下又有数个小类，如异物、擦花、脏污等。焊盘数据有五千多张，基材数据有八千九百多张。首先通过 Stable Diffusion，为二分类的数据集增加了两百多张。之后为多分类数据集的两个大类生成了两千多张。

## (三) YOLOv5 模型训练

下载 YOLOv5 后，在终端下载 requirement，使用 classify 文件夹中的 train.py 进行模型训练。参数方面，批量大小可以调整到占满显卡。迭代次数可以设置的较大，在训练过程发现模型收敛后可以手动停止训练。通过观察 train loss 和 valid loss 来判断模型是否在训练。而修改参数后，通过观察 loss 的下降程度和最后收敛后的 loss 值可以判断修改的参数是否有效。调整参数与文件路径后，选择 yolov5l-cls.pt 作为初始权重文件进行训练。

训练结束后，选择效果较好的模型作为结果模型。分别使用原来的数据集以及增强后的数据集进行训练，得到两个模型。

## (四) 模型推理

下载 inference 的文件后，调整文件路径，选择测试集的路径后，进行模型推理。因为最终目标是为了得到一个能够判断是否有缺陷的模型，所以测试集的数据只分为 Normal（正常）和 Abnormal（有缺陷）。最后得到正确率、召回率、漏

检率、过杀率和过滤率的数据。

### (五) 模型评估

完成模型的推理后，分析模型的推理结果。如果过于低于预期，需要重新检查训练步骤然后重新训练。最后，对比两个模型的推理结果。

## 四、问题及应对方法

### (一) 环境与文件相关问题及应对

首先，初次运行 YOLOv5 时，电脑内多个不同版本的 python 发生冲突导致了程序异常。解决方法是创建虚拟环境，以保证不同项目的环境相互隔离，不会互相干扰。

随后，在运行程序时，常常出现因文件出错导致的 bug。经过排除，发现是由于文件夹内的 DS\_Store 文件在文件内容变化时出错。解决方法是在运行前，先运行另一段检查文件的程序 check\_picture.py，检查文件夹内各个文件的有效性，将有问题的文件名打印在 check\_error.txt 的文件中。如果出现问题的文件是某些图片，则检查数据集并更换图片。如果出现问题的文件是 DS\_Store 文件，则更换文件夹。检查无误后再继续运行 python。如此可精准定位出错文件，节约排除出错故障的时间，提升效率。

### (二) Stable Diffusion 相关问题及应对

初次接触 Stable Diffusion 时，使用的是网上的 demo 版本，也就是在线进行有数量限制的生成。然而 demo 并不能进行大量的生成，不能满足填充数据集的需求，于是解决方法为下载 Stable Diffusion webui 版本进行本地部署，既能在本地运行 Stable Diffusion，也能通过网页 UI 进行更直观精准的调整。

为了能生成缺陷，刚开始进行的尝试是文生图的形式，也就是用 prompt 进行条件输入。然而，Stable Diffusion 的文生图并不能直接生成想要获得的工业缺陷图像，只会生成其他常见物体或人像。为了解决该问题，采用图生图的方法，用真实的缺陷图作为原图，引导模型生成更贴近真实缺陷的生成图，效果优于使用文生图。

然而，Stable Diffusion 有时会生成彩图，也就是噪声较多的未完成生成的图像。排查后发现是由于分辨率太低导致的，生成图像时未竟全功导致生成图像质量不佳。解决方法为调整分辨率直到能生成正常质量的图像，以保证后续大量生成的稳定性。

在生成过程中，Stable Diffusion 的不足之处逐渐显露。Stable Diffusion 最为常用的功能是人像绘画，也就是较为擅长画人。对于其他不泛用的领域，尤其是工业缺陷这种小众且需求具体的领域，Stable Diffusion 的效果十分有限，时常生成人脸，人的身形剪影等等。针对这种问题，采取了调整参数的策略。首先，调



整重绘幅度以限制 Stable Diffusion 的调整幅度，保证生成图依旧是缺陷图。将重绘幅度限制在 0.2-0.3 左右后，Stable Diffusion 能稳定生成缺陷图。其次，生成图有时会生成许多不够真实的缺陷图，比如类似油画的画风或生硬的缺陷边缘等缺点。为了提升真实度，解决方法是更换采样方法，如 DPM++ SDE Karras，虽然生成速度会逊色于 Euler a，但生成图的质量更具优势，缺陷更为真实。

### （三）模型训练相关问题及应对

模型训练中，模型正确率很低一直是个大问题。在数据集补充完毕后，模型在训练过程中表现较好，但推理结果不佳，正确率仅有 2%。原因出现在两个方面，一是数据集方面的问题，二是训练方面的问题。首先，生成图的数量并不足够，相比原数据集占比太小，很难对模型产生效果，而且生成图的类型分布并没有做到尽量填充数据集各类别，使之分布均衡。因此，调整了生成图的方案，在后来的多分类数据集中，生成数量至少为总数据集总数（原有数量+生成数量）的五分之一左右，且每个类别都涵盖，并且掌握数量，为较少的类生成的较多，直到数量与其他类型相持平，比如生成了两百多张可移动异物类型。

除此之外，训练参数也进行了调整。减少批量大小（batch size）以符合实际的硬件情况，合适的批量大小可以提高训练速度与模型收敛性能。降低迭代次数（epoch），掌握合适的提前结束训练的时机以防止过拟合。模型收敛的迹象通常通过观察损失函数来判断，以 train loss 和 valid loss 这两个损失函数的指标，来判断模型有没有在训练。如果没有正常双双下降，说明模型训练存在问题。Train loss 下降而 val loss 上升，说明模型可能过拟合。Train loss 和 val loss 都不变说明学习遇到瓶颈，需要调小学习率或批量大小。修改参数后，需要通过 loss 来判断修改参数有没有效。同时，更换初始权重文件，将 yolov5s 换成 yolov5l-cls 以获得更好的效果。

## 五、结论

### （一）结果

二分类的新模型正确率很低，两次训练后正确率分别在 2%和 22%左右，远不及原有模型的能力。

多分类的新模型对基材的效果有所下降，而对焊盘的效果有一点点提升。

分析原因，首先训练的步骤、操作与参数等仍需继续学习与调整。其次，Stable Diffusion 的生成图存在于原图较为相似的缺点。虽然可以改变形状和颜色，但改变不多且位置不会变化，大体上仍然较为相似，这对于模型的鲁棒性有所影响。最后，针对基材这个类，由于其中的许多缺陷存在难以识别的难点，生成较为困难，常常只能得到非常相似的缺陷或完全没有缺陷的图，因此基材的生成图质量并不佳。

## （二）未来改进

首先，在图像生成方面，可以继续更深入地学习扩散模型的相关原理框架、论文、代码等，利用工业真实缺陷的数据集去训练一个专门用于生成缺陷的模型，以生成效果更好的图。

其次，可以调整进行图像生成的具体操作。利用 Stable Diffusion 的局部重绘功能，通过局部调整，达到控制生成图缺陷大小位置的作用，以解决生成图与原图过于相似的问题。以及可以尝试结合文生图与图生图，引导 Stable Diffusion 生成更符合期望的图。亦可结合异常增殖<sup>[11]</sup>的方法，通过手动选定 Stable Diffusion 生成的缺陷达到位置变换的效果。

再者，与团队讨论后，发现可以通过为 Stable Diffusion 添加自动化的改进增加生成图片的效率。在生成过程中发现，Stable Diffusion 的图生图操作，是较为机械重复的。固定参数（如重绘幅度，采样方法等等）之后，可以设计 python 程序读取目标图片（图片全部重命名为数字），而后放入 Stable Diffusion 中，生成固定数量的生成图后放入指定的文件夹（output 文件夹）中，然后遍历直到最后一张图。如此，需要人工操作的只有挑图步骤，即挑选质量较好的生成图，提升了生成效率。然而，该方案的缺点在于无法实时根据图片情况调整参数以得到更好的图，对于某些需要调整参数才能生成出较好的图来说，仍然需要人工再次生成。但即使有这样的缺点，该方法仍能一定程度上加快生成周期，提高生成效率。

模型训练方面，可改进的地方是训练参数的调整和训练步骤的改进。可继续通过深入学习调整参数背后的原理，实际操作并验证对原理的理解，从而更好地训练出模型。

## 六、感想与总结

由衷感谢琶洲实验室（黄埔）提供我这次实习的机会，我感觉十分庆幸能在这里实习并在这样优秀的团队中学习成长。感谢徐博士和佳佳老师的指导、支持与激励，令我受益匪浅并增加了前进的动力。同时也要感谢办公室的各位同事前辈给予的帮助与教导，无私分享的这些宝贵经验与知识令我收获颇丰。以及感谢实习小组中的每一位成员，小组的合作精神和友好的讨论氛围促进了有力的进展。

这次实习的收获：

### （一）领略了请教与科研手段的重要性

古人云，敏而好学，不耻下问。下问尚且不耻，何况上问？当在实习中遇到困难或难点的时候，首先自己思考并尝试动手解决。当出现自己难以解决或花费很长时间依然毫无头绪的时候，应及时向前辈们请教，因为他们更有经验与实力去帮助解决看似十分困难的问题。比如之前在某个故障问题上操作了一上午依然没能解决，在请教了前辈后不出几分钟便迎刃而解。不但是积极请教，而且在办

公室也主动学习了他们的科研手段，包括查询检索相关论文的能力和如何进行高效工作的方法。没有这些科研手段，科学研究可能会举步维艰。所以，这些能力与方法都是进行科研的宝贵经验。

## （二）高端且实用的知识

在实习期间收获了深度学习、网络结构、模型训练等知识，从最根源的算法原理到应用方面的具体代码，我对深度学习和人工智能领域有了更深刻的理解，对于日后发展大有裨益。在其中的“大模型开发与赋能”专题讲习班中，不但了解了大模型相关的基本知识和发展，而且理解了其相关基本运作原理，更是对各种领域的大模型前沿研究有了一些了解。这些研究讲解详细且专业，虽然以目前的知识量并不能完全掌握这些非常顶尖的研究，但至少是一个好的开始。

## （三）知识的实际运用及经验

在学校中学习到的知识终究只是知识本身而已，需要经历“实战”，也就是实际的运用才能真正通透。实验室内学术研究与工程应用相结合的环境是最适合将知识快速落实到运用，我收获了许多学以致用经验。更重要的是，我充分理解了知难行易与知易行难的道理。关于知难行易，是在做之前想要充分理解其中的道理感觉十分困难，但实际上上手时却并没有想象中那么难。在使用 Stable Diffusion 前，我没有丝毫的相关经验，看到几百行的代码和大量的文件时更唯恐自己无法很好的掌握它的使用。然而实际上手后，很快就掌握了它的基本运用，并没有想象中那么困难。而所谓知易行难，在模型训练时有较强的体验。似乎是很简单且容易理解的微调，实际上却非常困难。我也明白了眼高手低断不可取，结合知识与应用才是硬道理，最终才能解决问题。

不但是知识的运用，我还自实习中掌握了许多技能。不但掌握了 Stable Diffusion 和 YOLOv5 模型的使用，而且理解了如何对大数据集进行预处理，这些技能更能帮助日后对于深度学习和计算机方面的学习研究。

更重要的是，我感受到了所学知识与实际工业场景之间的差距，以及图像生成方向的挑战与机会。在接触到工业运用之后，感受到了真实的工业运用时结构与算法的配合，结构更多的开始服务于算法。同时，也感受到图像生成这个方向的难度和机遇，虽然图像生成有些难学，但许多场景十分需要这方面的知识能力，因此挑战与机遇并存。

## （四）团队协作共同处理解决难题的重要性

在实习中，团队成员们合作无间，共同解决了实习方案梳理、环境文件部署、算法代码实现、图像生成改进和模型训练推理的各种挑战与难点。在相互协作中，形成了较为理想的合作氛围。

再次感谢每个人，我真诚地感激这段宝贵与难忘的实习经历。本次实习不但领悟了请教与科研手段的重要性，收获了知识及其运用，掌握了许多技能，感受到工业场景的实际性和图像生成方向的前景，更是提升了团队协作能力。

## 参考文献

- [1] Liu, Wei, et al. "SSD: Single shot multibox detector." *Computer Vision – ECCV 2016*, 17 Sept. 2016, pp. 21–37, [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, <https://doi.org/10.1109/cvpr.2016.91>.
- [3] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, <https://doi.org/10.1109/iccv.2017.324>.
- [4] Jin, Zhang, et al. "DWCA-Yolov5: An improve single shot detector for safety helmet detection." *Journal of Sensors*, vol. 2021, 2021, pp. 1–12, <https://doi.org/10.1155/2021/4746516>.
- [5] Wan, Dahang, et al. "Yolo-HR: Improved yolov5 for object detection in high-resolution optical remote sensing images." *Remote Sensing*, vol. 15, no. 3, 2023, p. 614, <https://doi.org/10.3390/rs15030614>.
- [6] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017, pp. 1137–1149, <https://doi.org/10.1109/tpami.2016.2577031>.
- [7] He, Kaiming, et al. "Mask R-CNN." *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, <https://doi.org/10.1109/iccv.2017.322>.
- [8] Zhong, BoLin, et al. "AC-Faster R-CNN: An improved detection architecture with high precision and sensitivity for abnormality in spine X-ray images." *Physics in Medicine & Biology*, 2023, <https://doi.org/10.1088/1361-6560/acf7a8>.
- [9] Goodfellow, Ian J, et al. "Generative Adversarial Networks" *Machine Learning (stat.ML)*, 10 Jun. 2014, <https://doi.org/10.48550/arXiv.1406.2661>.
- [10] Rombach, Robin, et al. "High-resolution image synthesis with Latent Diffusion Models." *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, <https://doi.org/10.1109/cvpr52688.2022.01042>.
- [11] Zhang, Hui, et al. "Prototypical residual networks for anomaly detection and localization." *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 4–5, <https://doi.org/10.1109/cvpr52729.2023.01562>.
- [12] Niu, Shuanlong, et al. "Region- and strength-controllable gan for defect generation and segmentation in industrial images." *IEEE Transactions on*

- Industrial Informatics*, vol. 18, no. 7, 2022, pp. 4531–4541,  
<https://doi.org/10.1109/tii.2021.3127188>.
- [13] Livernoche, Victor, et al. “On Diffusion Modeling for Anomaly Detection”  
*Machine Learning (cs.LG)*, 29 May. 2023,  
<https://doi.org/10.48550/arXiv.2305.18593>.
- [14] Wolleb, Julia, et al. “Diffusion Models for Medical Anomaly Detection”  
*Image and Video Processing (eess.IV)*, 08 Mar. 2022,  
<http://doi.org/10.48550/arXiv.2203.04306>.