

个人研究陈述 (Statement of Purpose)

宋有哲 (Youzhe Song)

[🏠 Homepage](#)

申请项目：计算机科学博士 (PhD in Computer Science)

申请地区：北美 (North America)

申请学期：2026 年秋季 (Fall 2026)

1. 引言：我的研究兴趣与思考

当前人工智能模型在各项任务中取得了显著的成功，但它们的内部工作原理很大程度上仍是一个“黑盒”。我们知道模型能做什么 (what)，并在一定程度上、一些细节中知道它是如何做到 (how) 的，但缺乏对于整个认知过程、思考过程符合人类直觉的理解。我认为，通往更通用、更可靠 AI 的重要一步，在于理解并设计模型内部的信息处理机制。我的研究兴趣聚焦于两个核心问题：第一，我们如何让模型像人一样，**理解并对齐来自不同来源（如图像、文本）的信息**？第二，我们能否在模型内部构建一个**符合人类直觉的、分步、有机的推理路径**？

我的目标是探索如何在高维的隐空间 (latent space) 中，实现这两种能力的融合，让模型不仅能“看懂”和“读懂”世界，更能在此基础上进行清晰、可靠的推理。

2. 研究实践一：设计更鲁棒的特征表示 (CoReFace)[1]

我的研究之旅始于对自监督对比学习的深刻向往。它不依赖昂贵的人工标注，而是从数据本身的结构中学习，这种优雅与潜力让我着迷。然而，现实的困境是，我在一个计算资源极为有限且每个学生研究方向都不一样的实验室中，环境迫使我放弃这一课题，并寻找一个更加实际的替代方向。

于是我发现了人脸识别这个完美的“沙盒”任务。我观察到它与对比学习在本质上的深刻相似性：两者都在高维空间中精心规划特征的分布。基于此，我提出了 CoReFace 框架，将对比学习代入到一般分类任务中，以实现对高维特征空间特征分布的优化。

在人脸这种细粒度识别任务中，传统的图像增强会破坏关键身份信息。无法从图像层面入手，我尝试思考对特征进行处理。我使用了 Dropout，它本质上是一种**特征层面的随机扰动**，它能在不破坏图像语义的前提下，为对比学习构成特征对。

进一步地，我发现并处理了分类、对比联合训练过程中的兼容问题，包括对比学习的监督信号过弱而无法产生有效损失，以及部分样本对的语义重复计算的问题。于是我深入优化了对比学习损失，加入了一个动态的相似度约束，在全训练过程中保持有效的训练梯度。对于样本语义重复的问题，我选择重组对比学习使用的样本对方案。

通过引入一个由对比学习引导的正则项，我主动地去“**排布**”特征空间的几何结构。最终，该方法成功将正负特征对的**相似度差异提升了 15%**。在这项工作中，我主导了从背景调研、设计方案到独立验证与论文撰写的完整生命周期，完成了从“学生”到“**学术**”的身份认知转变。

3. 研究实践二：构建异构数据的统一处理框架 (QGFace)[2]

完成 CoReFace 后，我渴望研究能与真实世界产生更直接的联系。这个想法在我翻看手机相册时受到了启发：相册自带的 AI 人脸聚类功能，在处理我的家庭合照时表现的不太好，经常漏掉某个人的照片。我意识到，现实世界中因拍摄距离、构图、光线等原因产生的低质量数据，与通用人脸识别常用的数据集有着重大的区别。

我希望以一种优雅、高效的方法，实现对不同质量数据的统一处理，于是 I 开始了 QGFace 项目。它的创作基于**单编码器 (single-encoder) 架构**结构，无需现有方案的超分、教师学生网络等等

冗余部分。清晰的图片便于判断人物，而十分模糊、有遮挡的图片其实无法判断它对应的身份。基于这种假设，我对不同质量的编码特征使用不同的监督信号：高质量数据使用分类损失，低质量数据使用对比损失，并用梯度截断来防止信息污染。

过程中我发现对比学习因样本对数量不足而效果不佳，于是我设计了一个**实时编码池**。它区别于对比学习常用的动量编码池，特征不来自动量编码器，而是来自 QGFace 的编码器，同时使用不同更新步下的分类向量的差更新这些编码结果。

为了追赶截稿日期，我全身心投入其中。那段时间，我的一天被切分成多个工作的时段，GPU 的工作时间才是我的休息时间。这段经历对我而言并非沉重的打击，那段时间我并没有陷入某种低沉消极之中，因为我知道自己正在创作有价值的内容：它让我确信，驱动我的不是外部的期望，而是解决难题时发自内心的追求和乐趣。

最终，QGFace 在几乎不牺牲高质量数据性能（**仅有 0.3%/95.5% 的微小性能权衡**）的情况下，在低质量数据集上达到了 SOTA，证明了我的研究哲学能指导兼具优雅与鲁棒性的实用方案。同时，该工作也是对单模态数据内容的一次挖掘，我始终认为当前方法对于模态本身的挖掘还有待提升，比如模型更偏好纹理而不是像人类依赖形状。挖掘单模态内部的对齐也将是我未来的工作重点。

4. 探索与聚焦：在实践中明确研究方向

尽管取得了一些学术成果，但投稿过程的反复失利与不确定性，让我陷入了一段巨大的迷茫。我开始强烈地自我怀疑：我真的适合学术的道路吗？为了寻找答案，我做出了一个决定：主动走进业界，对我内心真正的热情进行一次彻底的”探索”。

我先加入了由前 IBM 中国区 AI 总架构师方杰先生带领的初创公司 KeyoneAI。在这里，我将前沿的生成式 AI 技术落地为产品，感受着快速迭代的冲击，与团队协作与潜在用户交锋，做技术的布道者。不过这种与技术很近，与产品打磨焦灼，但是离思维的创新又有距离的感觉让我比较折磨。于是我后来离开了。

我想看看技术是如何赋能各种行业的，于是我加入了一家离 AI 源头更远的公司，世界教育者大会 (WWEC) 的组委会 [3]。作为唯一的技术负责人，我用技术赋能了一场上万人的国际化会议。在保障系统功能的同时促进团队数字化、开发内部工具支持 1000+ 复杂海报生成、链接 3000 观众、100 展商、10+ 供应商和各种临时会议，每周 80 小时工作近 3 个月。期间我使用 AI 编程功能生成了大量的新功能。我希望今后能构建一个学术自动化系统。

尽管这些都带来了巨大的挑战和经历后的成就感，但我发现，它们都无法替代我在科研过程中，发现在某个技术方向上科研人员几年一突破、一年几突破，并且让自己参与到这种突破中时，所感受到的那种纯粹的、智识上的兴奋与心流。这段”刻意疏离”的探索，让我获得了前所未有的清晰：我最深的渴望，是对事物底层原理的追问，是对现有方案的驳斥与重建，是旁征博引、有效创新的构建。

更重要的是，工作经历帮助我将曾经困扰我的过度敏感与自我审视和”内耗”，转变为我人生的沉淀。在工作过程中，我一直感受到内心的某种缺失，虽然我的工作内容是团队不可或缺的部分，但我内在的创造热情没有得到抒发。为了平衡心灵，我把自己带到户外，阅读身心灵相关的书籍。在所有这些过程中，关于实现智能的想法不时的就会闪现在我的脑海中。现在我更深刻地认为，**一个真正智能的系统，其标志不应是毫无瑕疵的单向推理，而恰恰是具备处理内在冲突、进行自我审视和迭代修正的能力**这正是人类”反思”与”纠结”的本质，是复杂推理与智能的基础。

5. 未来的研究蓝图：对齐语义，推理未来

现在，我带着更加清晰和坚定的目标来规划我的博士研究。我希望将我过去在**特征空间优化、异构数据处理**上的经验，与对**人类认知过程**的思考结合起来，专注于探索大模型的两个核心能力：

1. **构建统一的语义空间**：我的首要目标是研究如何将各种信息，有效映射到一个统一的、可解析语义的隐空间中。这不止是在强调跨模态的映射，也是对于单模态数据内容的挖掘。我相信，实

现高效的语义对齐，是构建能够全面理解世界的基础模型的第一步。我在 QGFace 项目中处理异构数据的经验，为我探索如何对齐和融合异构信息提供了宝贵的实践基础。

2. **在隐空间中实现结构化的推理过程**：在实现语义对齐的同时，我的另一个目标是隐空间中，设计并实现模型自主结构化的、类似”思维链”的推理路径。现有的模型类似于古早计算机使用纸带进行输入输出，而没有统一的大脑与思考过程。我希望模型不仅能得到答案，更能展现出一个可分解、可追溯的推理过程。这不仅能增强模型的可解释性和可靠性，也可能为解决更复杂的、需要多步逻辑的开放式问题提供新的途径。

总而言之，我的研究计划是双向并进的：一方面，通过**多模态对齐**让模型”看懂”更丰富的世界；另一方面，通过**潜在推理**让模型”想得”更清晰、更有逻辑。

6. 结论

我相信，凭借我在设计模型内部机制方面的实践经验，以及对构建”语义对齐”与”潜在推理”相结合的 AI 系统的清晰规划，我已准备好迎接博士阶段的挑战。我渴望在一个富有创造力和支持性的环境中，为构建下一代更强大、更可信的人工智能系统贡献自己的力量。

参考文献

References

- [1] Youzhe Song and Feng Wang. Coreface: Sample-guided contrastive regularization for deep face recognition. Pattern Recognition, 152:110483, 2024.
- [2] Youzhe Song and Feng Wang. Quality-guided joint training for mixed-quality face recognition. In 2024 IEEE International Conference on Automatic Face and Gesture Recognition (FG). IEEE, 2024.
- [3] Worldwide educators conference (wwec). <https://www.wwec820.com/>.