



## 附件一：2023 年 CCF-百度松果基金申报主题

### 目录

<b>1. 深度学习基础理论和技术</b> .....	<b>3</b>
1.1 优化算法和收敛性研究 .....	3
1.2 低精度训练和量化训练技术研究 .....	3
1.3 模型小型化技术研究 .....	3
1.4 深度强化学习技术 .....	3
<b>2. 深度学习框架技术</b> .....	<b>4</b>
2.1 支持动态 Shape 的中间表示设计 .....	4
2.2 神经网络编译器技术 .....	4
2.3 软硬协同的性能优化技术 .....	4
2.4 分布式训练技术 .....	5
2.5 深度学习框架扩展能力的研究 .....	5
2.6 新交互界面设计和下一代框架研究 .....	6
<b>3. 推理部署关键技术</b> .....	<b>6</b>
3.1 深度学习模型异构推理部署技术 .....	6
3.2 RISC-V 架构硬件适配技术 .....	6
3.3 RTOS 操作系统适配技术 .....	7
3.4 大语言模型推理加速技术 .....	7
3.5 基于编译器技术的推理加速技术 .....	7
<b>4. 大语言模型</b> .....	<b>7</b>
4.1 基础模型训练 .....	8
4.2 数据建设 .....	8
4.3 Prompt .....	8
4.4 基准测试 .....	9
<b>5. 计算机视觉</b> .....	<b>9</b>
5.1 图像分类关键技术 .....	9
5.2 3D/点云检测分割技术 .....	10
5.3 交互式图像分割关键技术 .....	10
5.4 OCR 关键技术 .....	10



<b>6. 跨模态大模型技术及 AIGC 应用</b> .....	<b>11</b>
6.1 大语言模型与视觉模型融合技术.....	11
6.2 跨模态多任务统一扩散模型.....	11
6.3 AIGC-文本生成图片技术.....	12
6.4 AIGC-文本生成视频技术.....	12
<b>7. AI 科学计算</b> .....	<b>12</b>
7.1 计算流体力学：复杂流动问题建模方法研究.....	12
7.2 分子动力学：AI 势函数模拟方法研究.....	13
7.3 地球科学：极端天气的智能预报研究.....	13
7.4 工程计算：流固耦合计算加速方法研究.....	13

## 1. 深度学习基础理论和技术

本课题期望通过对深度学习基础技术、基础理论层面的研究，来指导解决当前深度学习研发和应用上的挑战性问题，侧重对实际问题更有直接指导意义的或者对未来技术发展有引领意义的基础研究。

**建议研究方向：**

### 1.1 优化算法和收敛性研究

**概述：**研究能加速深度学习训练收敛的数值优化策略；研究大 batch size 训练时的收敛性问题，以支持更大节点数的分布式训练；研究大规模异步并行训练时的收敛性问题及优化策略，为多场景的异步训练效果提供理论指导。

**研究方法推荐：**从优化器、异步训练、模型实现、大 batch size 等不同策略角度，研究影响大模型训练收敛的重要因素；从端到端收敛时间角度，评估不同策略对大模型收敛效果的影响。

**验收标准：**给出不同算法策略对大模型训练收敛的理论分析和基于飞桨的实验验证结果，并在至少 3 个大模型上进行验证。

### 1.2 低精度训练和量化训练技术研究

**概述：**研究 FP16/FP8/BF16 等低精度浮点计算对训练收敛的影响；研究 INT8/INT16 量化训练的可行性，给出具备良好收敛效果的低精度训练方案，实现训练加速。

**研究方法推荐：**通过使用低精度运算，加速模型训练过程，自动判断模型中各个子网络结构对低精度的敏感性；对各个子网络使用低精度计算后对整体收敛性的影响给出评估，便于自动或人工决策对哪些子网络使用低精度运算能够在保证收敛的情况下对整体模型能够充分加速。

**验收标准：**给出全自动或半自动低精度训练方案，在 LLM 或其他常用大模型上，相比飞桨的自动混合精度训练 AMP 策略，模型收敛到相同指标的端到端训练时间缩短 10%以上。

### 1.3 模型小型化技术研究

**概述：**研究基于 LLM 大模型的信息冗余性的理论分析方法，以及包括稀疏化、低精度量化、蒸馏技术在内的模型压缩技术，保持模型精度的同时使模型体积更小、计算更快。

**研究方法推荐：**

研究一种理论分析方法，在特定数据和任务上，分析模型中每种结构的信息冗余性，用于指导模型压缩。针对生成式大语言模型，研究低成本的模型小型化技术，无需对模型进行重训练，即可达到无损压缩目的。

**验收标准：**基于飞桨在 2 个及以上开源 LLM 大模型上验证理论方法或技术的有效性，并支持在现有推理库上做部署加速。

### 1.4 深度强化学习技术

**概述：**强化学习指的是智能体通过和环境不断进行交互，根据反馈提升自己的决策能力，从而获得更高的奖励。强化学习在 LLM 大模型的训练过程中也发挥了关键作用。

**研究方法推荐：**

研究与 LLM 相结合的超大规模强化学习训练方法，针对策略训练与样本生成设计协同一体自适应的分布式训练策略。

**验收标准：**

采用新的训练方法，基于飞桨在 LLM 的 PPO 的强化学习分布式训练速度超越业界 SOTA 10% 以上。

## 2. 深度学习框架技术

本课题期望对深度学习框架技术进行系统、深入和前瞻的研究，解决当前深度学习框架中存在的 key 难点问题，探索下一代框架的设计实现。

**建议研究方向：**

### 2.1 支持动态 Shape 的中间表示设计

**概述：**中间表示 (Intermediate representation, IR) 是深度学习框架的核心数据结构，衔接前端和后端，对框架的功能完备性、性能优化便利性都非常关键。当前在基础算子拆分和神经网络编译器场景下，面临如何表示输入变量的 Shape 动态变化的问题。在基础算子拆分时，Shape 变化会影响算子拆分的语义，比如是否需要 broadcast；在编译器代码生成是，Shape 变化会影响代码生成和执行效率。

**研究方法推荐：**

提出一个更优的 IR 表示方案，能够解决 shape 不确定的问题，支持基础算子拆分的语义表达，支持神经网络编译器代码生成，避免重复构图和编译的开销。

**验收标准：**

基于飞桨的 IR 提供解决思路，在几个典型模型上跑通支持动态 Shape 的基础算子拆分和神经网络编译器流程，在大多数 Shape 设置下，对比固定 Shape 的速度下降不超过 5%。

### 2.2 神经网络编译器技术

**概述：**研究深度学习底层编译优化策略，包括运行时编译优化、压缩-编译协同设计、动态代码自动生成等，实现对多硬件的更好自适应支持、更高性能执行，研究编译器和手工优化结合的方法；研究和底层硬件协同的软硬一体整体优化方案，包括 AI 芯片和编译器的协同设计，在对应硬件上发挥出最佳性能。

**研究方法推荐：**

- (1) 研究新的自动/手动调优优化策略；
- (2) 研究针对 CINN 的 IR 的 Python 层 DSL 开发技术；
- (3) 研究编译器技术下的压缩、量化等技术；
- (4) 研究高性能的编译器动态 Shape 编译方案；
- (5) 研究深度学习编译器与新硬件的训练或推理适配技术。

**验收标准：**

- (1) 在相关算法上具有创新性、具有成为新式算法的潜力，并具有和前沿可比较的效果。
- (2) 基于飞桨工程化，在相关实验、测试、benchmark 上超越 SOTA 5% 以上，并能证明该效果由该研究带来，阐释解释其中的理论关系。

### 2.3 软硬协同的性能优化技术

**概述：**研究基于 CSR/C00 等通用格式的非结构化稀疏或硬件结构化稀疏等技术，以实现训练、推理加速；研究稀疏算子加速技术及在典型场景的应用；研究典型模型在通用计算硬件上的性能优化技术。

**研究方法推荐：**

- (1) 使用稀疏算法对 GPT 类 LLM 或其他大模型进行训练加速，精度指标与非稀疏化持平，可以考虑使用算法对权重、输入数据等进行高度稀疏化，配合高效稀疏 OP 优化进行整体加速。
- (2) 结合硬件编程模型和特征，针对重要模型结构或算子，重新设计并行策略和实现方案（如 Flash-Attention）、并进行极致优化，达到更快的计算速度。

**验收标准：**

- (1) 基于稀疏技术的模型端到端训练时长相比非稀疏算法减少 15%以上。
- (2) 基于飞桨重要计算单元的性能优化在代表性模型上性能提升超过 SOTA 实现 10%。

## 2.4 分布式训练技术

**概述：**研究更高效的分布式训练策略，包括训练效率的提升和扩展性的提升；研究图神经网络模型的高效大规模训练方法；研究低配网络条件下的并行训练算法，减小或甚至消除网络带宽限制对训练速度和收敛效果的影响；研究异构硬件/集群下的新的分布式训练方案；研究大规模参数服务器架构技术。

**研究方法推荐：**

- (1) 研究超大规模模型并行训练优化技术，通过并行策略的改进提高模型训练加速比，并保持收敛性。
- (2) 研究低带宽网络条件下的梯度压缩算法，保证模型收敛的同时提高分布式并行效率。
- (3) 研究超大规模图学习参数服务器引擎，支持多机多卡训练和多级存储。
- (4) 研究自动并行相关技术，给定模型和集群资源，在满足显存约束情况下，设计高效搜索算法自动地给出模型的最优并行策略。
- (5) 研究无状态化 Collective、参数服务器训练和推理表征方法，支持资源弹性调度和故障重提。

**验收标准：**

- (1) 基于飞桨提出一种分布式训练优化策略方法，在千卡规模集群上性能提升 20%以上。
- (2) 基于飞桨提出一种低带宽网络条件下的梯度压缩算法，分布式训练性能提升 80%以上，兼顾模型收敛和训练性能。
- (3) 基于飞桨提出一种超大规模图训练的图分割算法，并完成多机多卡训练和多级存储功能实现，收敛效果优于业界最优实现。
- (4) 在 GPT 等经典模型上，基于飞桨自动搜索出来的并行策略性能不低于人工设定的最优并行策略，同时搜索时间在小时以内。支持 Collective 和参数服务器训练和推理过程的弹性调度和自动容错，包含故障节点和 hang 节点自动检测、无中断式的训练推理节点自动替换和扩缩容功能。

## 2.5 深度学习框架扩展能力的研究

**概述：**研究如何在保证用户数据不泄露的情况下，更高效地在飞桨上实现深度学习模型训练和部署的方案。研究在深度学习开发、训练、部署等环节增加哪些措施，可以保证用户的模型数据安全。

**研究方法推荐：**

联邦学习，通过分布式模型训练的方式，有效帮助多个机构在满足用户隐私保护、数据安全和政府法规的要求下，进行数据使用和机器学习建模，解决数据孤岛和数据隐私的问题。数据加密，通过加密算法和加密密钥将明文转变为密文，而解密则是通过解密算法和解密密钥将密文恢复为明文；在深度学习场景下，是否能找到合适的加密解密算法，使得

模型可以在加密的数据上训练，在不涉及训练数据明文的情况下，拿到模型的预测结果。

**验收标准：**

形成切实可行的技术方案，基于飞桨在一些典型场景和任务上实现原型系统开发和验证。

## 2.6 新交互界面设计和下一代框架研究

**概述：**研究命令式编程和声明式编程更好的结合方式，做到灵活性和效率、部署兼顾；探索新的编程范式、用户交互方式以及对应的底层实现逻辑，进一步提升用户开发效率、简化底层架构复杂度；研究框架和平台层更好的结合融合方式，探索提升用户体验的全新产品形态。

**研究方法推荐：**

- (1) 针对推理的场景而言，动转静后的代码应该执行效率高且易于部署。此时应该更多用编译的思路理解命令式编程界面，提供足够的编译反馈让用户改进原始的命令式代码，并承诺经过编译后的代码易于部署且高效。
- (2) 针对训练的场景而言，动转静的完备性和高效性同等重要。用户代码的动态性必须得到保障，只有涉及到静态图的部分才做性能优化。研究过程中需要秉承准静态假设，将代码中的可静态部分清晰地剥离开，将其编译成静态图，而剩下的部分继续维持动态运行。

**验收标准：**

- (1) 针对推理场景而言，好的动转静功能应该尽可能少的、精准地给用户提供编译反馈信息，方便它改进，同时，编译后的代码应该尽可能高效。
- (2) 针对训练场景而言，好的动转静功能应该覆盖足够的 Python 语法规则，维持动静的语义一致性，再尽可能地高效。
- (3) 基于飞桨完成原型系统开发和方法有效性验证。

## 3. 推理部署关键技术

本课题拟研究推理部署关键技术，包括深度学习模型异构推理部署、RISC-V 架构适配、RTOS 操作系统适配技术、大语言模型推理部署技术等。

**建议研究方向：**

### 3.1 深度学习模型异构推理部署技术

**概述：**研究单模型在异构硬件上的调度方案，实现模型推理的延迟或是吞吐的提升；基于 NVIDIA GPU Jetson 硬件的异构部署，如 X86 加 NVIDIA GPU 的组合，NVIDIA GPU Jetson 系列内部 CPU 加 GPU 加 DLA 的组合；手机上 ARM CPU 加 GPU 加 DSP 加 NPU 的组合等。

**研究方法推荐：**

通过图优化、硬件信息自动调优 AutoTune 方法寻找最优性能的异构计算的组合。

**验收标准：**

Paddle Lite 可以基于 ARM CPU 加 OpenCL 异构组合，通过调度优化，性能优于现有默认方案 20%，CPU 和 GPU 负载增加不能高于现有方案 10%。

### 3.2 RISC-V 架构硬件适配技术

**概述：**RISC-V 是与 X86/ARM 齐平的开源硬件架构，随着国内外 RISC-V 产业联盟以及深度学习的发展，越来越多的硬件厂商设计了基于 RISC-V 架构的深度学习硬件。研究 Paddle Lite 对 RISC-V 架构硬件的适配技术，支撑应用在 RISC-V 架构的硬件上高效部署。

**研究方法推荐：**



在 Paddle Lite 新增 RISC-V 后端，参考 ARM CPU 并基于 RISC-V SIMD 指令开发 RISC-V 算子内核 Kernel。

**验收标准：**

实现 Paddle Lite 对 RISC-V 的支持，并且能够满足实际生产的应用需求。第一阶段，完成 25 个算子实现，完成 ResNet50、MobileNetV1/V2/V3、SSD 模型的适配；第二阶段，完成 110 个算子实现，完成 FastDeploy 超过 70 个模型的适配。

### 3.3 RTOS 操作系统适配技术

**概述：**随着深度学习的发展，逐渐开始有应用将深度学习的模型基于 RTOS 系统进行部署。本研究项目希望能够完成 Paddle Lite 对 ARM CPU 在 RTOS 系统的支持。

**研究方法推荐：**

修改编译脚本、方法以及框架代码，完成 RTOS 操作系统的适配。

**验收标准：**

实现 Paddle Lite 对 RTOS 的支持，并且能够满足实际生产的应用需求。完成 VxWorks、FreeRTOS 任意一款 RTOS 操作系统的适配，跑通 ResNet50 分类模型示例，并将代码合入到 Paddle Lite 仓库。

### 3.4 大语言模型推理加速技术

**概述：**LLM 模型推理性能对落地及成本至关重要，由于显存瓶颈通常需要多卡甚至多机进行部署，研究 LLM 模型性能加速方案，包括多卡并行切分、通信优化、Transformer 高效推理、量化及稀疏推理加速，或其他 LLM 模型推理加速算法，节省时延或提升吞吐。

**研究方法推荐：**

并行方法及通信优化：在多卡推理中，通信往往一定的占比，可研究多卡并行方法 或 通信优化；时延优化：可优化 Context 计算 及 Decoder 解码截断的更高效的 Transformer 实现 或 其他高性能算子；吞吐优化：可通过显存优化(如 CacheKV 量化等)提升 Batch Size，或通过其他手段提升整体吞吐。

**验收标准：**

至少选择 1 个 PaddleNLP 中 GPT、BLOOM、LLaMA、GLM 及 ChatGLM 等大模型，基于飞桨实现时延或吞吐较已有方案提升 20%以上。

### 3.5 基于编译器技术的推理加速技术

**概述：**编译器技术已被广泛应用于深度学习框架之中，研究编译器技术加速模型 GPU 推理，尤其 LLM 模型推理，不限于 FP16、BF16、INT8、INT4 等精度类型。

**研究方法推荐：**

- (1) 基于飞桨编译器打通 LLM 模型推理；
- (2) 基于飞桨编译器 CINN 研究 LLM 模型在 GPU 上的 FP16、BF16 高效推理实现，如高性能 Kernel 生成、算子融合等；
- (3) 基于飞桨编译器 CINN 研究 LLM 模型在 GPU 上的 INT8、INT4 高效推理实。

**验收标准：**

基于飞桨 CINN 编译器跑通 GPT、BLOOM、LLaMA、GLM 等至少 1 个 LLM 模型，性能达到非编译器手工优化的 90%以上。

## 4. 大语言模型

近年来，大语言模型（Large Language Model）在自然语言处理领域取得了巨大的成功。大语言模型是指使用大量文本数据训练的深度学习模型，可以生成自然语言文本或理解语言文本的含义。大语言模型可以处理多种自然语言任务，如写作、问答、对话等，是通向人工智能的一条重要途径。

**建议研究方向：**

#### 4.1 基础模型训练

**概述：**研究大语言模型的模型结构、预训练方法、微调算法、人工反馈学习等技术。

**研究方法推荐：**

- (1) 新型大语言模型结构：研究新型大语言模型结构，使其具备超强长文本建模能力和语义表征能力，在学习效果和效率方面领先 Transformer；
- (2) 大模型推理溯源与提升：研究大模型推理能力来源，通过思维链构建、课程学习等方式，提升大模型在复杂任务的推理能力；
- (3) 模型提示学习：研究基于超小样本的 Prompt-tuning 技术，无需对模型进行重训练或者仅添加少量参数，即可在下游任务上取得更优效果；
- (4) 多模态预训练：研究多模态统一表征学习与融合，使得通过不容模态学习，增强视觉模态、语言模型及跨模态能力；

**验收标准：**基于飞桨在 20 个以上典型应用任务上，效果有显著提升，并具备实际应用价值；并支持现有大模型云平台应用。

#### 4.2 数据建设

**概述：**数据建设在大语言模型的指令微调和对齐训练中发挥着至关重要的作用，直接影响着大语言模型是否能够有效地激发能力，且与人类的价值观保持一致。大语言模型的数据构建存在众多值得研究的问题和方向，包括但不限于数据体系构建和应用、数据质量检查、数据分布调优、数据组织形式探索、数据缺陷发现等。通过对数据的调整和优化，我们希望能够更高效的训练出效果更好的大模型。

**研究方法推荐：**

- (1) 数据体系的建设及应用：研究如何建设全面的数据体系，以及围绕数据体系的扩充数据广度和深度，可以围绕某一大类任务如文本创作，或围绕一个学科如数学；
- (2) 数据质量检查：研究大模型数据质量 CheckList 及相关工具；
- (3) 数据分布优化：研究混合数据的最佳比例和分布，以及如何进行自动的数据分布调优；
- (4) 数据缺陷发现：研究如何测试和发现模型及对应数据中的缺陷，并针对性的进行解决。

**验收标准：**

- (1) 基于以上研究内容，形成一套能够自动/半自动的执行的方法，并可以有效解决对应的问题。
- (2) 基于以上研究内容，实现对数据的优化，对应的评估集合上效果有显著提升，G:B 提升 10%以上。

#### 4.3 Prompt

**概述：**研究面向大语言模型的 Prompt 设计、自动逻辑控制和外部工具调用机制对大模型的发展和具有重要价值。Prompt 设计需要编写准确、明确的指令，以引导模型产生期望的输出。自动逻辑控制则关注如何通过自然语言的描述来实现模型的自主决策和流程控制。此外，外部工具调用机制的研究可以使模型与其他实用工具进行无缝集成，例如知识



查询、API 调用等，以提供更丰富、准确的信息。这些研究旨在提高大语言模型的应用能力和性能，为各领域的专业用户提供更加高效、可信赖的解决方案。

**研究方法推荐：**

- (1) 智能 Prompt 引擎架构：研究基于大型语言模型的 Prompt 自动化生成机制，以及 Prompt 模板和用户需求的自动匹配机制。这种方法旨在最大化利用 Prompt 提升大模型生成效果，节约人设计 Prompt 的成本。
- (2) 自适应逻辑控制：研究基于大模型的目标规划、任务拆解能力；从数据驱动的角度，提升模型自适应逻辑控制能力、自我决策能力等。
- (3) 外部工具调用：研究如何以 Prompt 机制实现大语言模型和外部工具接口的交互。包括接口协议设计，Prompt 设计、调优，zero-shot 工具调用能力验证等。

**验收标准：**

- (1) 基于飞桨建设一套 Prompt 引擎架构，支持 Prompt 自动匹配、生成，提升大模型效果 10%以上。
- (2) 大模型初步具备目标规划、任务拆解、任务调度/决策能力。
- (3) 基于一版大模型，实现至少 5 个外部工具的调用能力，正确率达到初步可用状态 (70%以上)。

#### 4.4 基准测试

**概述：**设计大语言模型评测体系，建设有效的测试任务、方法、指标和工具，客观、全面和高效地评测大语言模型。

**研究方法推荐：**

- (1) 构建能深度评测大语言模型能力水平的评测体系、评测任务及测试用例集。
- (2) 建设自动评测方法及工具集，能够基于评测任务及测试用例集对不同模型的效果进行自动化评估，产出可信的定量指标对比结果。

**验收标准：**

- (1) 能够全面衡量模型能力，包含不同难度等级的评测任务及测试用例集。
- (2) 基于飞桨实现的自动化评估流程及工具集，能够对比不同模型的效果，产出客观可信的指标对比结论。

### 5. 计算机视觉

本课题希望对深度学习技术在计算视觉中的前沿问题和产业化实践中的实际难题进行分析和探索。在前沿问题探索中，着重方法的创新性、领先性和可推广性，鼓励探索基础模型的突破，填补视觉技术领域的空白。在实际产业化难题解答中，针对实际场景，着重方法的效果和效率，促进产业化的高效、高质量发展。

**建议研究方向：**

#### 5.1 图像分类关键技术

**概述：**图像分类是计算机视觉中最基础的任务，是整个计算机视觉的根基，研究人员在图像分类上的探索的方法往往可以迁移到计算机视觉的其他任务中。该领域的研究主要分围以下方面：对服务器端和移动端等不同设备，在推理速度受限的情况下，系统化研发更高精度的图像分类网络结构和预训练模型，包括自监督等前沿技术；在特定的数据集上，探索基于已有的预训练权重的模型迁移能力。

**研究方法推荐：**

- (1) 骨干网络结构设计：研发在同等速度下精度更高的骨干网络，适配骨干网络到其他视觉任务中；
- (2) 自监督学习：通过对比学习、掩码学习等方案，学习出图像的通用特征，其权重用于其他视觉任务的预训练权重；
- (3) 半监督学习：结合有标签数据和无标签的数据，提升目标数据集的精度；
- (4) 知识蒸馏：通过学习教师模型的知识，提升学生模型的精度；
- (5) 迁移学习：基于给定的预训练权重，提升在目标数据集的精度。

**验收标准：**

- (1) 基于飞桨研发出更高效的网络结构，在特定设备上的时延更小，精度更高。
- (2) 基于特定的网络结构，研发出精度更高的预训练模型，且可以在其他视觉任务有较好的迁移效果。
- (3) 加载特定的预训练模型，在具体的垂类任务上，研发出更高的精度的模型。

## 5.2 3D/点云检测分割技术

**概述：**以机器人导航、混合现实、智能交通、无人驾驶等场景为基础，研究三维点云识别、点云生成、点云分割、实例分割、3D 检测、3D 跟踪等相关领域领先技术。

**研究方法推荐：**

- (1) 基于各种传感器（相机、激光雷达、毫米波雷达）获得的不同模态的数据，进行各种阶段各种方式的融合；
- (2) 研究设计一种集成多任务，包括但不限于感知、定位、跟踪、决策与规划的，端到端的深度学习神经网络；
- (3) 自动挖掘实际场景中的难例数据，并能基于此类数据完成对已有系统的提升。

**验收标准：**

- (1) 基于飞桨研究 3D 算法，在 nuScenes、Waymo 等公开数据集上，精度优于 SOTA；
- (2) 在 Orin、T4 主流推理芯片上，推理性能达到实时或优于 SOTA 方案。

## 5.3 交互式图像分割关键技术

**概述：**图像分割作为计算机视觉三大基础任务之一，在各个领域具有广泛的应用需求。图像分割取得快速发展，但是依旧面临一些挑战，比如图像分割标注困难，标注成本比分类、检测高出数十倍。

**研究方法推荐：**

研究人员提出交互式分割、弱监督分割、主动学习等技术，希望系统化研究该方向，进一步提升图像分割标注效率。

- (1) 结合大模型能力，使用文本 Prompt 进行交互式分割。
- (2) 挖掘可解释性和弱监督分割的结合。

**验收标准：**

- (1) 基于飞桨完成研究，在对应领域公开数据集上实现 SOTA 效果。
- (2) 在 PaddleSeg 分割套件、EISeg 标注工具中，初步完成落地验证。

## 5.4 OCR 关键技术

**概述：**OCR 作为计算机视觉的基础任务之一，在跨模态、大模型等众多任务中起到重要支撑作用。期望研究人员可深入探索视觉信息和语义信息的深度融合方法，研发如何利用语义信息辅助文字检测和文字识别新策略；研发轻量、高效的表格识别策略；鼓励探索训练方式的突破，借助蒸馏、半监督等技术手段，提供高效模型训练方案。同时针对训练数据难获取难标注的现状，探索有效的文字类数据生成算法。

### 研究方法推荐:

- (1) 结合语义信息的文本识别策略
- (2) 基于分割或基于序列识别方法的表格识别策略
- (3) 针对大量无标签数据的自监督、半监督学习训练策略
- (4) 高效的知识蒸馏、迁移学习策略
- (5) 基于文生图的文字类数据生成策略

### 验收标准:

- (1) 基于飞桨, 完成文本识别、文本检测、端到端和表格识别等方向的 SOTA 算法, 显著提升现有模型精度。
- (2) 基于飞桨, 完成知识蒸馏、自监督、半监督、迁移学习的方案, 提供小成本模型训练方案。
- (3) 基于飞桨, 打造文字类的数据生成工具, 基于生成数据提升文本识别或文本检测效果。

## 6. 跨模态大模型技术及 AIGC 应用

**概述:** 跨模态是指在不同类型的数据和模态之间进行交互和融合的技术。跨模态大模型技术创新带来了 AIGC 应用突破, 让 AI 在预训练过程中同时学习模态间和模态内的多种关联性, 提升“图像”, “视频”和“文本”跨模态语义匹配效果; 以扩散模型为中心统一多任务架构; 技术驱动应用创新包括文生图、图生图、文生视频等。

### 建议研究方向:

### 6.1 大语言模型与视觉模型融合技术

**概述:** 跨模态大模型融合了大语言模型和视觉模型的特点, 并且展现出更多样的能力, 例如视觉文档、图像摘要、图像推理等。跨模态大模型拥有更大的参数量和计算复杂度, 研究低成本模型训练方案, 充分利用大语言模型和视觉模型已有能力, 通过结合更多微调策略及下游任务, 高效提升跨模态大模型效果。

### 研究方法推荐:

- (1) 参考 BLIP2、miniGPT4、VisualGLM 等基础模型, 研究低成本跨模态大模型预训练方案;
- (2) 参考 InstructBlip 方法, 研究基于 Instruction Tuning 的跨模态大模型微调策略;
- (3) 参考 Grounding、Open-Set Segmentation 方法, 研究图生文预训练融合更多类型视觉任务。

### 验收标准:

- (1) 基于飞桨, 提出低成本跨模态大模型预训练方案, 可灵活融合不同规模大语言模型及视觉模型。
- (2) 基于飞桨, 开发跨模态大模型微调策略, 在各类下游任务中达到 SOTA 效果。
- (3) 基于飞桨, 实现图生文与其他视觉任务统一的预训练方案。

### 6.2 跨模态多任务统一扩散模型

**概述:** 当前跨模态大模型领域主要包含三大方向: 图生文预训练, 跨模态视觉任务和文生图。文生图领域底座以扩散模型为主, 该方法能够有效扩展到其他跨模态方向, 研究跨模态领域多任务统一方案, 实现更加灵活的图片文本输入、图片文本输出能力。

### 研究方法推荐:

- (1) 参考Unidiffuser, Versatile Diffusion 模型, 研究基于飞桨实现文生图、图生文统一扩散模型;
- (2) 参考UViT 基础模型, 研究基于Transformer 架构实现更高效的扩散模型底座。

#### 验收标准:

- (1) 基于飞桨, 提出统一扩散模型方案, 多任务统一模型效果相比单任务效果有提升
- (2) 基于飞桨, 提出更高效的扩散模型底座, 文生图效果性能优于Stable Diffusion。

### 6.3 AIGC-文本生成图片技术

**概述:** 扩散模型技术突破以及Stable Diffusion 的开源使得文本到图像生成迎来了爆发实现破圈。在此基础之上, 如何进一步提高文生图的质量、增强生成的可控, 并更好的满足图像编辑、小样本定制化等多种场景的需求具有很强的研究价值与实际意义。

#### 研究方法推荐:

研究基于如人类反馈的文生图训练方法, 更高效的模型架构, 如Subject Driven 的可控文图生成技术, 如Text-Guidance->Image-Guidance 图片编辑技术, 高分辨率文生图, 如多个主体及其组合定制化, tuning-free 的小样本个性化文生图技术。

**验收标准:** 基于飞桨, 在公开权威评测集上, 验证所提出的方法与业界SOTA 在图片生成的定量评估指标和定性评估效果和训练推理速度上的有效性。

### 6.4 AIGC-文本生成视频技术

**概述:** 在文生图扩散模型取得技术突破和巨大进步后, 文生视频是AIGC 下一个亟待突破的应用方向。如何结合当前文生图模型的技术方法, 扩展应用到视频生成领域, 针对低资源和大规模不同场景, 实现视频生成质量的显著提升, 具有很强的研究价值与实际意义。

**研究方法推荐:** 研究Zero-Shot、Few-Shot 及大规模训练的文生视频模型, 提升生成视频的质量, 质量包括但不限于主体一致性、动作连贯性、分辨率、视频长度等。

**验收标准:** 基于飞桨, 在公开权威评测集上, 验证所提出的方法与业界SOTA 在视频生成质量上定量评估指标、定性评估效果和训练推理速度上的有效性。

## 7. AI 科学计算

AI 科学计算是指使用人工智能方法、利用计算机再现、预测和发现客观世界运动规律和演化特征的全过程。通过AI 学习自然规律、求解数学模型并应用于工程实践和科学探索, 解决航空航天、船舶制造、生物计算、地球科学、能源勘探等领域的难题。本课题期望探索AI 技术和科学计算任务相结合的创新方法, 具备良好的实用性和可推广性, 更高效解决各领域的科学计算问题。

#### 建议研究方向:

### 7.1 计算流体力学: 复杂流动问题建模方法研究

**概述:** 使用数值方法在计算机中对流体力学的控制方程进行求解, 从而可预测流场的流动, 用AI 的方法更好地解决实际场景中多相流、流固耦合、湍流模拟、超高速流场仿真等复杂流动问题。例如, 心脏数字孪生技术(Digital Twin) 结合医学影像、固体力学、流体力学等多学科知识, 模拟心脏的复杂运动和射血功能, 构建心脏数字孪生需要解决复杂的流固耦合(Fluid Structure Interaction, FSI) 问题。

#### 研究方法推荐:

- (1) 开发高效贝叶斯优化算法和PINN 模型, 实现参数快速反演;

- (2) 研究神经网络代理模型技术，实现个体化流固耦合分析；
- (3) 利用对抗学习和主动学习的方法，选取具有代表性和困难的样本进行物理场仿真。

#### 验收标准：

- (1) 完成基于飞桨平台的数字孪生算法，分析速度对比传统算法有明显提升。
- (2) 发表人工智能顶级会议不少于 3 篇，高水平期刊不少于 2 篇。

### 7.2 分子动力学：AI 势函数模拟方法研究

**概述：**分子动力学通过使用计算机来模拟分子、原子体系的运动，从而以动态观点考察系统随时间演化的行为，用 AI 方法更好地解决材料属性预测、新材料发现、新能源材料等问题。研究 AI 势函数模拟方法，并使用 LAMMPS 或 GROMACS 等工具，实现分子动力学模拟。

#### 研究方法推荐：

- (1) 基于飞桨和 DeepMD-kit 工具，实现如 DeepPotential-Net 方法从量子力学数据中学习和预测势能曲面；
- (2) 基于飞桨实现如 SchNet、SphereNet 等基于 GNN 的势函数模拟方法。

#### 验收标准：

- (1) 基于 GPU、飞桨、DeepMD-kit 和 LAMMPS 跑通分子动力学模拟全流程。
- (2) 基于飞桨实现 DeepPotential-Net、SchNet、SphereNet 等势函数模拟方法。

### 7.3 地球科学：极端天气的智能预报研究

**概述：**开展全球智能预报、区域智能预报、非结构化观测数据分析等问题的模型研究、算法设计与工程优化相关探索，以提升极端天气事件（暴雨、高温、台风等）预报精度，可以最大限度降低极端事件对生命安全、国家利益带来的危害。现有天气预报大模型以 ERA5 作为初始场，与传统模式预报对比存在初始场精度不一致，导致对比结果不公平、难以嵌入实际预报过程的问题，因此需要结合实际模式预报初始场设计思路，结合实际观测数据来源及精度，深入探索再分析场与传统初始场的差异，开辟与传统模式预报对齐的赛道。

#### 研究方法推荐：

- (1) 探索智能预报模型时序建模，在 3D 预报框架基础上，嵌入 SwinTransformer 或 RNN，提升智能预报的有效预报时长。
- (2) 针对现有智能预报算法对极端事件捕捉不足的缺陷，从模型结构设计、损失函数设计角度，提升大模型对极端事件预报的精度，并从理论上论证。

#### 验收标准：

- (1) 在飞桨上实现的大模型框架及相关代码；
- (2) 发表相关领域具有影响力的论文 1 篇。

### 7.4 工程计算：流固耦合计算加速方法研究

**概述：**流固耦合问题广泛存在于各种科学和工程领域，例如航空航天、机械、土木、海洋工程、汽车、能源、生物医学等。通过研究这些流固耦合问题，我们可以深入了解相关系统的机理，从而为进一步的优化或控制提供理论指导。如何利用 AI 加速流固耦合计算成为一个非常值得研究的方向，能促进相关领域的研究发展，并为工程设计和优化提供准确高效的指导。

#### 研究方法推荐：

- (1) 科学计算加速：基于飞桨平台开发物理信息约束神经网络模型，完全求解或者部份求解流固耦合问题，对比利用传统方式求解有加速作用。



- (2) 优化工程设计：基于飞桨开发工程设计 Encoder-Decoder 神经网络模型，将低维的设计参数空间投影到高维的参数空间，减少搜索空间的大小，从而更快的找到最优设计参数。

**验收标准：**

- (1) 选择一些具体的实际工程和科学问题，将基于飞桨开发的数值计算方法应用于这些案例研究中。通过与常规方法的比较，评估所提出方法的性能和优势，并分析其在实际应用中的潜在价值和局限性。
- (2) 模型的预测结果与精确数值模拟结果相比较，应具有合理的准确性和精度。
- (3) 该方法应该能够显著提高流固耦合数值模拟的计算效率，减少计算时间和资源消耗。
- (4) 该方法应具有一定的通用性，能够适用于不同类型的流固耦合问题。