arXiv:2305.05216v1 [physics.flu-dyn] 9 May 2023

文章信息

Dat用于深度学习应用的参数化 U形弯管流动数据集

作者
Jens Decke*[1], Olaf W¨unsch[2], Bernhard Sick[1]

所属机构
智能嵌入式系统，卡塞尔大学，Wilhelmsh¨oher Allee 73，D-34121 卡塞尔[1]流体动力学，卡塞尔大学，M¨onchebergstraße 7，D-34125 卡塞尔[2]

通讯作者的电子邮箱地址和 Twitter 账号
jens.decke@uni-kassel.de

关键词
机器学习、计算流体动力学、OpenFOAM、设计优化、形状优化、多物理场、共轭传热

摘要
该数据集包含 10,000 个 U形弯管形状中的流体流动与传热仿真。每个仿真均由 28 个设计参数描述，并借助计算流体动力学方法进行处理。该数据集为研究设计优化领域的各种问题与方法提供了全面的基准。针对这些研究，可以采用监督式、半监督式和无监督式深度学习方法。该数据集的一项独特之处在于，每种形状都可以通过三种不同的数据类型表示：设计参数与目标值的组合、几何形状和数值模拟解变量的五种不同分辨率的二维图像，以及基于数值网格单元值的表示。第三种表示方式使深度学习方法能够考虑数值模拟的特定数据结构。本研究还发布了用于生成这些数据的源代码和容器。

规格表

主题	计算力学
具体主题 area	流体流动和传热现象（共轭 传热）的数值建模，采用计算流体动力学（CFD）
数据类型	表格 图像 代码文件（Python、Bash）
How the data 已获取	使用 Python、OpenFOAM 和其他开源软件。数据 是在该系自己的计算集群上生成的。所使用的代码 与数据集一同公开提供。
数据格式	Raw 筛选后

描述 of 数据收集	该数据集由 10,000 个 U形弯管形状样本组成。每个样本由 28 个参数描述。 这些样本均已生成、评估并 使用 Python 和 OpenFOAM 进行处理。每个样本均独立 且服从相同分布。每种形状的性能由 其压力损失和冷却能力决定。
数据源位置-tion	•机构：卡塞尔大学，智能嵌入式 系统系 •城市：卡塞尔 •国家：德国
Data 可访问性-ity	存储库名称：卡塞尔大学数据存储库 DaKS 数据标识符：10.48662/daks-17 数据直接网址：https://daks.uni-kassel.de/handle/123456789/50 代码标识符：10.5281/zenodo.7717020 代码直接网址：https://zenodo.org/badge/latestdoi/612247257 提交时包含的条目
相关研究 文章	Decke, J., Schmeißing, J., Botache, D., Bieshaar, M., Sick, B., Gruhl, C.（2022）。NDNET：用于异常和新颖性检测的统一框架—— tion. 载于：Schulz, M., Trinitis, C., Papadopoulou, N., Pionteck, T.（编者）计算机系统架构。ARCS 2022。《计算机系统讲义》Com-计算机科学，第 13642 卷。Springer，Cham。 https://doi.org/10.1007/978-3-031-21867-5 13 _

数据价值

- 尽管该多物理场问题十分复杂，但仍可从流体动力学的角度对其进行解读和理解。
- 该数据集对从事拓扑优化、形状优化或一般部件设计优化的研究人员很有价值。
- 这些数据可用于开发新的设计优化算法。尤其是深度学习方法需要大量数据，而目前这类数据尚未公开。
- 偏微分方程的数量庞大，因此目标变量也很多，这些目标变量对于迁移学习相关方法尤其有研究价值。例如，有可能减少实际需要求解的偏微分方程数量。
- 数据集的每个样本都以三种不同方式表示（参数、图像、单元）。

目标

该数据集可能会引起结合深度学习与基于数值模拟的设计优化领域的研究人员的兴趣。在这一领域中，该数据集可作为基准，因为它用途广泛，每个样本都由三种不同的数据类型表示。采用监督学习、无监督学习和/或半监督学习方法，开展多模态特征、迁移学习、主动学习以及基于图像和/或图数据的学习研究，都是可行的。该数据集最初用作案例研究，用以展示所开发框架在异常检测和新颖性检测方面的广泛适用性 [1]。为便于其他研究人员使用该数据集并开展后续研究，该数据集以 CC-BY-NC-4.0 许可协议公开提供。

数据描述

这些数据包含了用于优化 U形弯管设计的流体流动和传热仿真结果。所介绍数据集的结构如图1所示，具体说明如下。

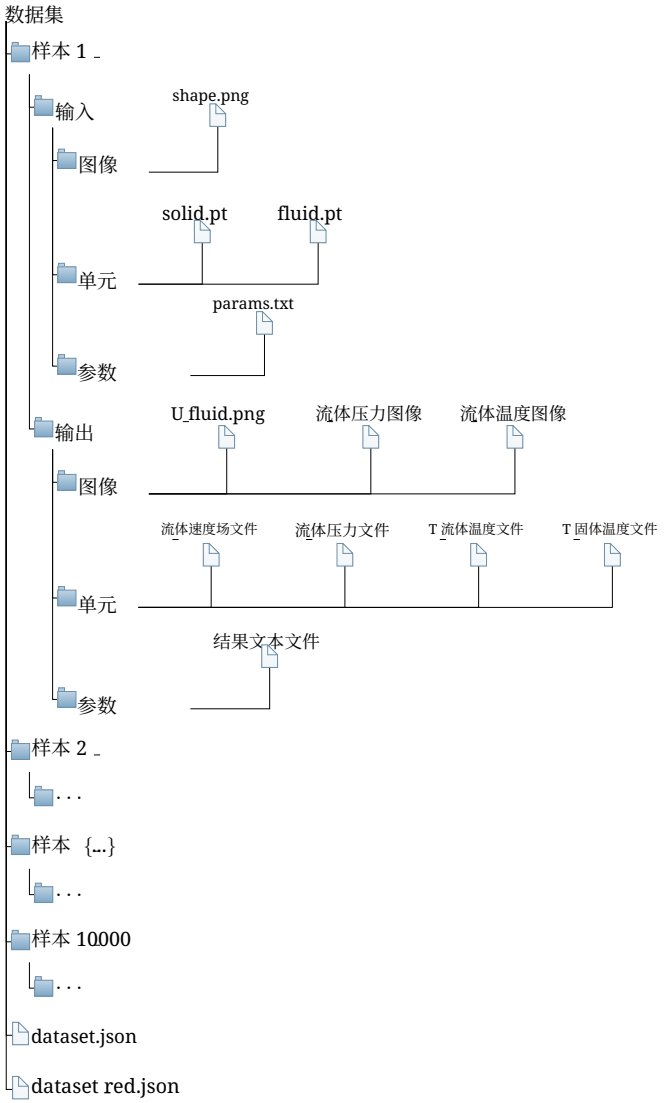


图1：当前数据集的结构。每个独立同分布（i.i.d.）样本都有自己的文件夹，其中包含输入和输出目录，用于存储三种不同表示形式各自对应的文件。数据集的结构信息见$dataset.json$ 和$dataset\ red.json$。

该数据集包含 10,000 个相互独立且服从相同分布（i.i.d.）的设计。每个设计在主路径下都有一个专属文件夹，其中包含一个输入文件夹和一个输出文件夹。这些文件夹中还有其他目录，存放不同设计数据格式的数据。这三种数据格式分别为定义设计的参数值、设计图像以及数值网格。参数值的表示形式存储在 Parameters 文件夹中。Input 中包含一个 params.txt 文件，其中存有一个包含 28 个设计参数的向量，部分参数的取值范围为 -1 至 1，部分为 0 至 1。Output 中包含 results.txt 文件，其中存有一个由两个目标值组成的向量：一是入口与出口之间以 $[Pa]$ 为单位的压力损失，二是以 $[K^2m^2]$ 为单位表示设计冷却性能的性能指标。第二种数据表示形式为 .png 图像。输入中包含 shape.png 文件，用于表示流体区域和固体区域。输出中则为微分方程的解变量各提供一张图像。这些解变量为速度矢量 U、压力 p 和温度 T。文件名与之对应（U fluid.png、p fluid.png、T fluid.png）。所有图像均提供五种不同分辨率。为简洁起见，图1中未展示全部图像。第三种设计表示方式是使用为计算设计而生成的数值网格。这些数据位于 Cells 文件夹中，并以 PyTorch [2]张量格式保存。输入提供了 solid.pt 和 fluid.pt 文件，其中包含相关信息

关于数值网格中相应单元中心的坐标。输出包含微分方程中每个解变量对应的一个文件，其中存储了单元中心处的解（*U fluid.pt*、*p fluid.pt*、*T fluid.pt*、*T solid.pt*）。此外，数据集的主路径中还有一个*dataset.json* 文件，其中包含数据集中所有文件夹和文件的数据结构。另有一个*dataset red.json* 文件，其中存储了所有成功模拟样本的结构信息。根据所需的研究内容，使用其中一个*json* 文件可能更为有利。

生成这些数据的源代码功能已在“实验设计”一节中详细说明。文件 *main.py* 是主文件，负责启动数据生成。通过文件*export batch file.py* 可触发一个新进程，并由此生成一个样本。文件*export blockmesh file.py* 是 Python 与 OpenFOAM 之间的接口。*read folder.py* 是一个实用函数，用于生成 **.json* 文件。文件*Allrun template.sh* 用于向 Podman 容器传递命令。

实验设计、材料与方法

本节首先介绍设计空间参数化、材料属性、边界条件和两个不同的目标函数，最后说明实验装置。

设计空间参数化：本研究开发的参数化模型基于冯·卡门流体动力学研究所的一个基准测试案例。该案例最初由 Verstraete 等人提出。 [3]。Goeke 和 W¨unsch [4]扩展了原始模型，加入了更多设计参数和一个固体区域。图2所示为将原始三维问题简化为二维后的结果。

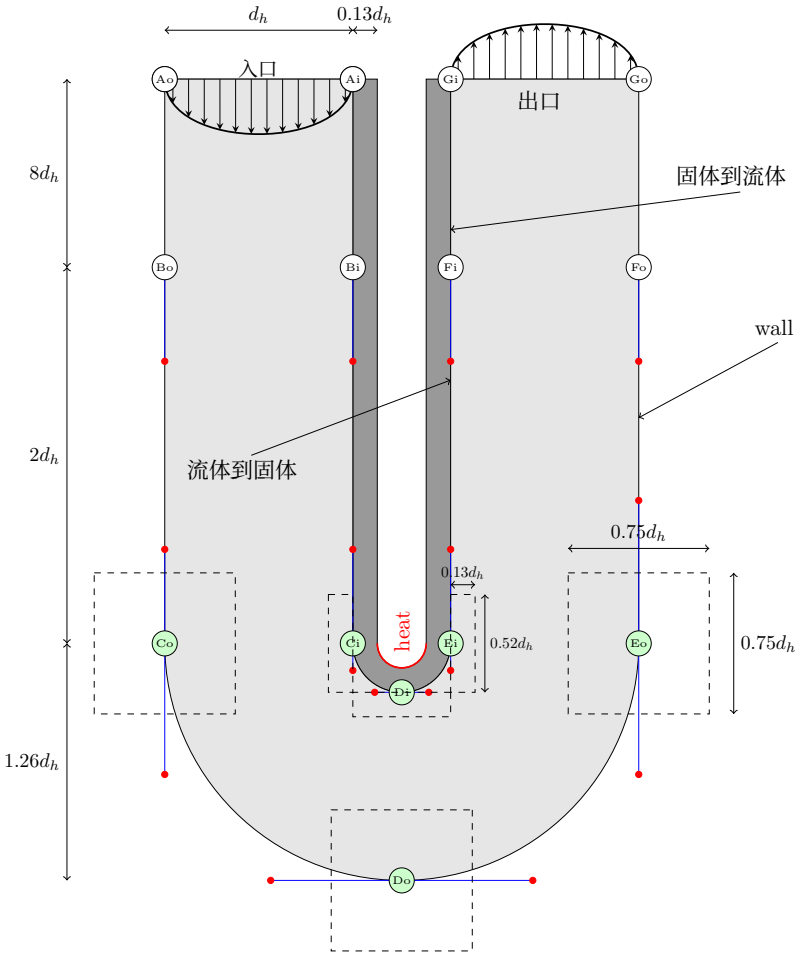


图2：参数化后的 初始几何形状及其边界点和曲线参数。改编自 [1]

它具有一个圆形U形弯管，外半径为 $1.26 \cdot d_h$，流体区域的边界由内半径

$0.26 \cdot d_h$，并连接了厚度为 $0.13 \cdot d_h$ 的固体。入口和出口截面的长度为 $10 \cdot d_h$。所用水力直径为 $d_h = 0.075m$。图 2展示了初始 U形弯管，其中包含固体（深灰色）和流体（浅灰色）两个区域。U形弯管被划分为 A 至 G 不同层级。这些层级表示所在层的位置 o （外层）或 i （内层）。白色填充圆圈标记的边界点是固定的，因此整个系统没有自由度。然而，绿色填充圆圈标记的边界点代表两个设计参数，因此为整个系统贡献两个自由度。为避免显示入口和出口通道的完整长度，图2中将 A/G 层级至 B/F 层级的距离压缩显示。绿色填充圆圈表示几何形状的边界点，可在虚线框内变化。这些框表示相应边界点的设计参数范围。外层框的尺寸为 $0.75d_h$ x $0.75d_h$ 和 $0.52d_h$ x $0.13d_h$ ，内层框的尺寸为。外层上每个绿色填充圆圈标记的点，其 x 和 y 坐标均在 -1 到 1 之间参数化，原点（初始解）对应的参数值为 (0,0)。正参数值会扩大流通面积，而负值则会使流通面积收窄。内层上，一个自由度也在 -1 到 1 之间参数化，另一分量则在 0 到 1 之间参数化。使用三阶 B´ézier 曲线连接边界点。每条曲线由两个曲线设计参数控制，如图2中的红点所示。蓝线标示了每个曲线参数的原点。红点及其对应的蓝线只能沿单一轴（x 或 y）移动。每个曲线参数值均在 0 到 1 之间，值为 0 时，曲线参数恰好位于相应的边界点上。六个边界点和八条曲线（每条曲线有两个设计参数）总共产生 28 个设计参数。在下图 3中展示了两个设计示例。

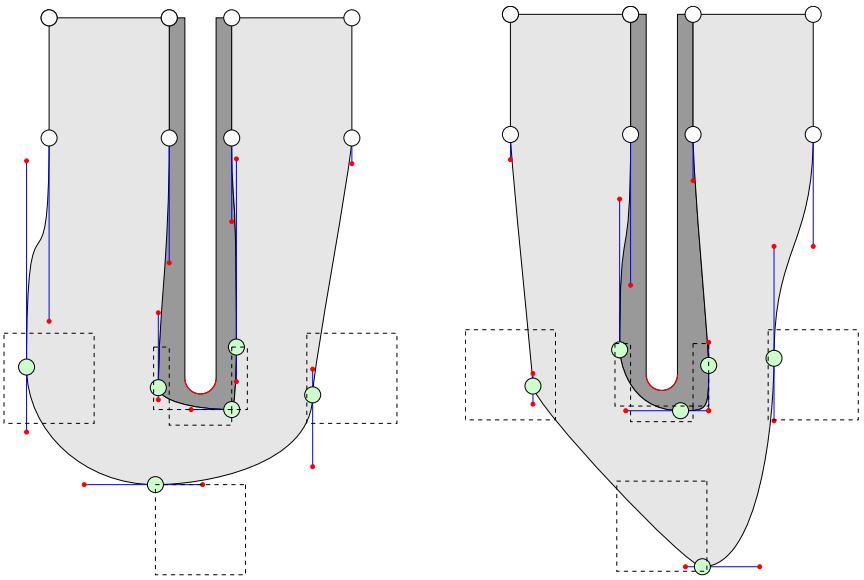


图3：两个具有不同设计参数的样本

材料属性和边界条件： 以下介绍基本材料属性和所用边界条件。对于数据集中的每个样本，这些参数均保持不变。图2所示的U形弯管由流体区域和固体区域组成。所选流体为空气，固体区域由结构钢制成。假定所用材料属性为常数，汇总于表格2。

表2：固体区域（钢）和流体区域（空气）的材料属性

材料	热导率	热容	密度	普朗特数	运动黏度
	$\lambda[W/(mK)]$	$c_p[J/(kgK)]$	$\rho[kg/m^3]$	Pr[-]	$\nu[m^2/s]$
固体	25	490	7850	-	-
流体	0.0262	1004.5	1.2	0.7	$1.5 \cdot 10^{-5}$

各区域的边界条件分别指定，并按照OpenFOAM中常用的名称进行描述 [5]。 $noSlip$ 条件适用于所有壁面上的流动。这意味着流速必须为零。入口入口处指定固定值。由于为自由流动，出口出口 处速度的法向梯度必须为零。流体以300K的温度进入。压力p_{rgh} 为静压 p ，在入口处计算，并扣除静水压力。绝热壁面壁面和极限壁处使用固定通量压力边界条件。出口处的压力p_{rgh} 指定为固定值零。各边界处均计算静压 p 。在表面热热处，给定特定的热通量外壁热通量温度为 $q = 100,\ 000W/m^2$ 。图2 展示了所有边界；流体区域的详细说明见表格 3 ，固体区域的详细说明见表格4。

表3：流体的边界条件

	入口	**出口**	**流体到固体**	**wall**
速度 **v**	固定值	零梯度	固定值	无滑移
值	值列表	均匀 (0 0 0)	均匀 (0 0 0)	均匀 (0 0 0)
温度 T	固定值	零梯度	湍流温度- 耦合挡板混合	零梯度
值	均匀 300	均匀 300	均匀 300	均匀 0
压力 p_{rgh}	固定通量压力	固定值	固定通量压力	零梯度
值	均匀 0	均匀 0	均匀 0	均匀 0
压力 p	计算值	计算值	计算值	计算值
值	均匀 0	均匀 0	均匀 0	均匀 0

在流体与固体的界面处，温度必须相同，因此，提供给流体的热流必须与固体传导的热量相对应。使用

湍流温度耦合挡板混合，即可实现这两个边界条件，并且必须为两个区域分别指定。因此，需要定义两个边界，而不是一个边界。对于流体区域，该界面定义为流体到固体。固体区域则相反，定义为固体到流体。

表4：固体的边界条件

	heat	**wall**	**固体到流体**
温度 T	外壁热通量 通量温度	零梯度	湍流温度 耦合挡板混合
值	均匀 300	均匀 300	均匀 300
压力 p	计算值	计算值	计算值
值	均匀分布 0	均匀分布 0	均匀 0

0.05 的马赫数允许采用 不可压缩 假设。雷诺数为 40000 表明流动完全处于湍流状态。为考虑湍流，决定采用双方程 $k-\omega\ SST$ 模型。湍动能的湍流边界条件 k、湍流黏度ν_t 和比湍流耗散率 ω 在表格 5中列出。

表5：湍流边界条件

	入口	**出口**	**wall**	**前后面**
湍动能 k	固定值	入口出口	kqRWallFunction	空白
值	值列表	内部场	内部场	
湍流黏度 ν_t	固定值	计算值	nutk壁面函数	空
值	值列表	内部场	内部场	
特定湍流 耗散率 ω	固定值	入口出口	ω壁面函数	空白
值	值列表	内部场	内部场	

边界条件入口出口 通常与零梯度相同。不过，当速度矢量指向边界附近的域内时，它会变为固定值 。此固定值 是入口处的值。

为确保入口处形成充分发展的速度分布，预先计算了一个一维槽道流。此外，湍流参数 k、 ν_t 和 ω 取自此预计算结果。这确保了二维U形弯管边界具有合适的初始解。

基于初始几何形状进行了全面的网格研究。采用流体区域横截面上有 60 个单元、固体区域有 15 个单元、沿流动方向有 780 个单元的网格，得到了最可靠的结果。借助 Coletti 等人通过粒子图像测速获得的测量数据，以及 Hayek 等人的三维模拟，对初始解进行了匹配和验证。 [6] [7]为节省计算时间，后续生成样本时以初始问题收敛后的数值解作为起点。

目标函数：引入两个目标函数，用于评估不同设计并衡量其性能。流体力学中一个显而易见的优化问题是使压力损失最小化。压力损失

$$J_1 = \frac{1}{A_{in}} \int_{A_{in}} p_{in} \, \mathrm{d}A_{in} - \frac{1}{A_{out}} \int_{A_{out}} p_{out} \, \mathrm{d}A_{out} \tag{1}$$

通过入口和出口压力在面积 A 上的积分之差 p 计算，并除以面积 A进行归一化。较低的压力损失伴随着较低的泵送功。一个重要方面是保持体积流量的边界条件不变，即泵送介质的量保持恒定，因为流量为零会导致压力损失为零。

第二个目标函数用于量化冷却性能，这借助加热表面的温度来实现。在热流已确定的情况下，较低的加热表面温度意味着更大的传热。目标函数

$$J_2 = \int_{A_{heat}} (T_{heat} - T_{in})^2 \, \mathrm{d}A_{heat} \tag{2}$$

由加热壁温度 T_{heat} 与入口温度 T_{in} 之间的温差的平方构成，并在加热表面 A_{heat}上进行积分。该目标函数由Goeke和W¨unsch于 [8]提出并采用。从物理角度来看，这两个目标相互排斥，因为不存在能够同时使二者达到最小值的设计。其原因在于，对流传热和压力损失分别受到雷诺数、流速和壁面剪切应力的耦合影响。 J_1 会倾向于较低的雷诺数。然而，由于较高的雷诺数有利于增强对流传热，这将降低目标 J_2。

实验设计： 计算机实验的工作流程如图 4所示。实验搭建过程中仅使用了免费开源软件。所用编程语言为 Python [9], , 工作负载管理器 Slurm [10], , 容器软件 Podman [11] （用于半虚拟化 Linux 操作系统）以及 OpenFOAM [5] 作为 CFD 软件。

该 ①**Design-Sampler** 会生成一个包含 28 个相互独立且同分布的随机变量的向量。在此步骤中，这些变量会进一步处理为 OpenFOAM 可读取的格式，以创建网格。随后，
使用 ②**作业调度器** 对预定义数量的作业进行并行化。作业调度器会为每个样本创建一个作业，该作业完全自动化，能够在计算集群的多个节点间调度作业。仿真完成后，由 ④**数据收集器**收集结果，并持续调度新作业，直至生成所需数量的样本。

图4展示了模拟过程，其中包括三个大型平行的深灰色方框。借助 Podman 半虚拟化软件，每次模拟都会创建一个容器。该容器运行基于 Linux 的客户操作系统、CFD 软件 OpenFOAM，以及处理数据所需的其他辅助工具。由 *Design-Sampler* 提供的文件会被 ③.1 **Mesh-Generator** 接收，用于生成计算网格。随后，在 ③.2 **网格控制**中检查网格质量。由于数值模拟的解质量对网格质量非常敏感，

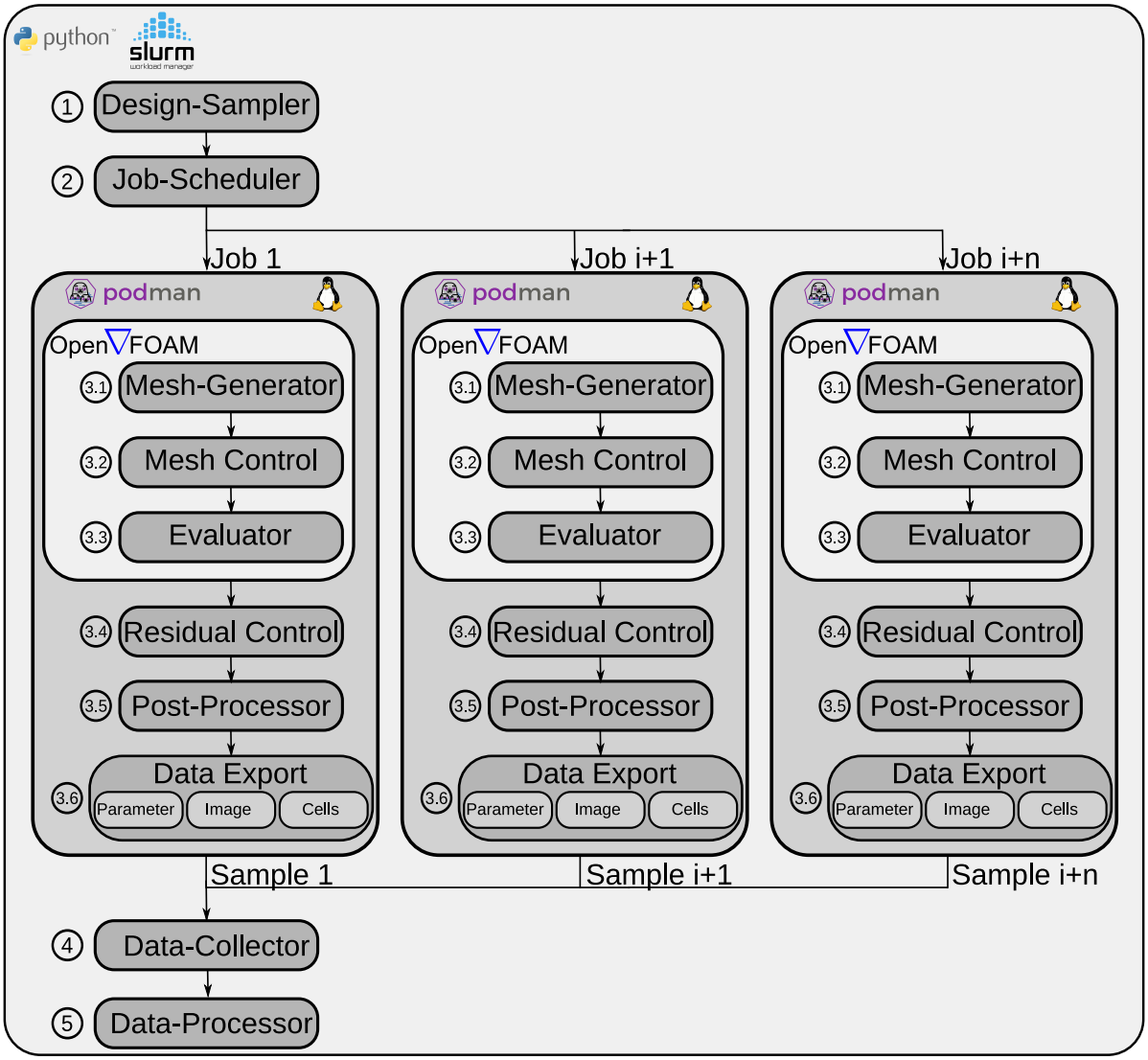


图4：生成 U形弯管设计数据集的实验设计

因此，不满足既定质量准则的网格样本会被弃用。此外，参数化可能会导致内层和外层相交，这在物理上是不可能的，因此不允许这种情况发生。用于评估网格质量准则的参数包括网格中单个单元的最大偏斜度、最大纵横比和最大非正交度。

使用 *chtMultiRegionSimpleFoam* 求解器作为 ③.③ **评估器**。它是一种稳态求解器，用于求解浮力驱动的湍流和固体导热，并处理固体和流体区域之间的共轭传热。
该 ③.④ **残差控制**使用各个数值解变量的残差来监控解的质量。监控残差有两个目的：首先，在达到特定解质量后立即停止仿真；其次，在仿真完成后，筛除不满足最低解质量要求的样本。使用 AMD EPYC 7002 处理器（加速时钟频率最高可达 3.35GHz）对单个样本进行仿真，耗时为 15 至 240 分钟。网格质量或解质量不足的样本会被赋予人为设定的高目标值作为惩罚。是否应将这些样本纳入后续研究或予以排除，取决于所研究的问题，因此它们原则上仍属于该数据集的一部分。该 ③.⑤ **后处理器**根据数值解计算已定义的目标值，并将其传递给 ③.⑥ **数据导出**模块。随后，该模块按照图1所示，根据演示的数据格式对解进行整理和导出，并将其转发给 ④**数据收集器**。最后一步是，

该 ⑤**数据处理器**会删除不必要的日志文件。系统会生成一个 *.json 文件，用于表示数据集现有的文件夹和数据结构。该 *.json 文件可让数据快速、高效地用于深度学习任务。

伦理声明
本研究未涉及人体或动物实验。

CRediT 作者贡献声明
Jens Decke：概念构思、方法学、软件、调查、数据整理、撰写初稿、可视化。**Olaf W¨unsch**：审阅与编辑、监督、资源提供。**Bernhard Sick**：资源提供、项目管理、经费获取。

致谢
第一作者谨向 Stephan Goeke 致以诚挚谢意，感谢他在整个项目过程中始终如一的支持、有意义的讨论和宝贵建议。他出色的合作以及乐于分享想法的态度，对本研究的形成起到了至关重要的作用。

特别感谢智能嵌入式系统系的服务器团队，尤其是 Dr. Christian Gruhl 和 David Meier。他们为创建该数据集，对计算集群的部分功能进行了定制。

本研究由联邦经济事务与气候行动部（BMWK）在“基于人工智能的电机拓扑优化（KITE）”项目（19I21034C）框架内资助。

利益冲突声明
作者声明，他们不存在已知的竞争性经济利益或可能影响本文所报告工作的个人关系。

参考文献

[1] J. Decke、J. Schmeißing、D. Botache、M. Bieshaar、B. Sick 和 C. Gruhl，《NDNET：异常与新颖性检测统一框架》，载于 计算机系统架构 （M. Schulz、C. Trinitis、N. Papadopoulou 和 T. Pionteck 编），（Cham），第 197–210 页，施普林格国际出版公司，2022 年。

[2] A. Paszke、S. Gross、F. Massa、A. Lerer、J. Bradbury、G. Chanan、T. Killeen、Z. Lin、N. Gimelshein、L. Antiga、A. Desmaison、A. Kopf、E. Yang、Z. DeVito、M. Raison、A. Tejani、S. Chilamkurthy、B. Steiner、L. Fang、J. Bai 和 S. Chintala，“PyTorch：一种命令式风格的高性能深度学习库”，载于神经信息处理系统进展 *32*，第 8024–8035 页，Curran Associates, Inc.，2019年。

[3] T. Verstraete、F. Coletti、J.Bulle、T. Vanderwielen 和 T. Arts，“内部冷却通道中压力损失最小化的U形弯管优化——第一部分：数值方法”，《涡轮机械杂志》，第 135 卷，2013年 6 月。

[4] S. Goeke 和 O. W¨unsch，《基于伴随方法的弯曲风道拓扑优化》，《应用数学与力学论文集》，第19卷，第1期，e201900438页，2019年。

[5] C. Greenshields，《OpenFOAM v10 用户指南》。英国伦敦：OpenFOAM基金会，2022。

[6] F. Coletti、T. Verstraete、J. Bulle、T. Van der Wielen、N. Van den Berge 和 T. Arts，《内部冷却通道中U形弯管的最小压力损失优化——第二部分：实验验证》，《涡轮机械杂志》，第135卷，2013年6月。

[7] M. E. Hayek、Q. Wang 和 G. M. Laskowski，《针对U形弯管流动的雷诺平均纳维–斯托克斯涡黏性模型的伴随优化》，载于*2018*年*AIAA*航空航天科学会议，（佛罗里达州基西米），美国航空航天学会，2018年1月。

[8] S. Goeke 和 O. W¨unsch，“基于伴随法的共轭传热系统拓扑优化”，《应用数学与力学论文集》，第 17 卷，第 771–772 页，2017 年 12 月。

[9] G. Van Rossum 和 F. L. Drake Jr，《*Python* 教程》。荷兰阿姆斯特丹数学与计算机科学中心，1995 年。[10] A. B. Yoo、M. A. Jette 和 M. Grondona，“Slurm：用于资源管理的简单 Linux 工具”，载于《并行处理的作业调度策略》，2003 年。[11] H. Gantikow、S. Walter 和 C. Reich，“用于 HPC 的 Podman 无根容器”，载于《高性能计算 》（H. Jagode、H. Anzt、G. Juckeland 和 H. Ltaief 编），（Cham），第 343–354 页，施普林格国际出版公司，2020 年。