

표면 곡률 기반 샘플링을 통한 Transformer 기반 3차원 유동장 예측 정확도 개선

유병걸^{*†}, 양선웅^{**}

*한양대학교 융합기계공학과, **한양대학교 ERICA 기계공학과

Improving Transformer-Based Three Dimensional Flow Field Prediction via Surface Curvature-guided Sampling

Byeonggeol Yu^{*†}, Sunwoong Yang ^{**}

^{*}Dept. of Mechanical Convergence Eng, Hanyang Univ., ^{**} Dept. of Mechanical Eng, Hanyang Univ. ERICA

Key Words: Optimal design, Anchored-Branched Universal Physics Transformer (AB-UPT), Computational Fluid Dynamics (CFD), Transformer, Computation mechanics

Abstract

High-resolution automotive flow-field prediction requires efficient representation of large surface and volume meshes while preserving local geometric features that influence the flow. In the original Anchored-Branched Universal Physics Transformer (AB-UPT), surface anchors are selected randomly, so local variations in vehicle surface geometry are not explicitly considered during anchor selection. This study introduces a geometry-informed surface-anchor sampling method to examine whether anchor placement based on local surface variation improves flow-field prediction. Local surface variation is estimated from the covariance eigenvalues of neighboring surface coordinates and combined with a uniform sampling component. The proposed method is compared with random sampling under the same model architecture, training conditions, and fixed surface-query evaluation protocol using the subsampled DrivAerML dataset.

1. 서 론

전산유체역학(computational fluid dynamics, CFD)은 차량 형상에 따른 표면 및 체적 유동장을 평가하는 데 활용된다. 고충실도 CFD는 복잡한 형상과 주변 유동을 해상하기 위해 많은 격자를 사용하여, 여러 형상을 반복적으로 평가할 때 계산 비용이 증가한다. DrivAerML은 차량 형상과 고충실도 CFD 결과를 제공하는 데이터셋으로, 형상과 유동장의 관계를 학습하는 대리모델 연구에 활용되고 있다.⁽¹⁾ 이러한 대리모델에서는 제한된 계산 자원으로 대규모 공간 정보를 처리하면서 물리량의 예측 정확도를 유지하는 방법이 요구된다.

Alkin 등⁽²⁾은 이를 완화하기 위해 anchored-branched universal physics transformer(AB-UPT)를 제안하였다. AB-UPT는 전체 mesh 중 일부 point를 anchor로 선택하여 정보를 처리하고, query point가 이를 참조하도록 구성하며, geometry, surface, volume branch를 통해 차량 형상과 유동장을 표현한다.

자동차 외부유동 CFD에서는 차체 표면의 형상 변화가 유동의 국부적인 가속과 감속, 압력구배, 경계층 발달 및 박리 위치에 직접적인 영향을 준다. 특히 곡률 변화가 큰 표면에서는 표면 법선과 유동 방향이 빠르게 변하면서 pressure와 wall Shear stress와 같은 물리량에 대해 공간적 구배가 커질 수 있으며, 이러한 국부 분포는 항력과 양력과 같은 적분 공력계수에도 영향을 준다. 따라서 고충

실도 CFD에서는 형상 변화와 유동구배가 큰 영역을 충분한 공간 해상도로 표현하는 것이 해석 정확도 측면에서 중요하다.

그러나 기존 AB-UPT의 surface anchor는 차량 표면에서 무작위로 선택되므로, 제한된 수의 anchor가 형상 변화가 큰 영역을 충분히 포함한다는 보장이 없다. 본 연구에서는 이러한 CFD 해석상의 공간 해상도 문제를 surrogate model의 sampling 관점에서 다루기 위해, 각 표면점의 국부 곡률을 surface anchor 선정 기준에 반영하였다. 인접점 분포로부터 국부 형상 변화 정도를 계산하고, 곡률이 큰 영역의 표면점이 anchor로 선택될 확률을 높이되, 특정 영역으로의 편중을 제한하기 위해 무작위 선택 성분을 함께 적용하였다. 이후 기존 random sampling과 동일한 데이터 및 학습·평가 조건에서 비교하여, 형상 변화가 큰 표면 영역에 계산 표현을 집중하는 방식이 물리량에 대한 유동장 예측 오차에 미치는 영향을 정량적으로 평가하였다.

2. 이론 및 실험

2.1 AB-UPT와 표면 앵커의 역할

AB-UPT의 앵커는 attention의 key와 value를 제공하는 공간적 참조점이며, 쿼리는 그 정보를 참조하는 예측 위치이다. 전체 토큰 수를 N_t , 앵커 수를 M 이라 하면, 앵커 사이의 연산과 쿼리-앵커 사이의 연산은 각각 M^2 및 $(N_t-M)M$ 에 비례한다⁽²⁾. 본 연구는 이 attention 구조를 변경하지 않고 표면

[†] Presenting Author, ybk2684@hanyang.ac.kr

앵커를 선택하는 규칙을 비교한다. 곡률 점수는 별도의 입력 채널이나 손실항으로 추가하지 않으며, 형상 분기와 체적 앵커 샘플링 및 분기별 디코더는 공통으로 유지한다.

비교 기준은 Alkin 등⁽²⁾이 제안한 AB-UPT의 무작위 surface anchor sampling을 바탕으로, 본 연구의 공통 평가 조건에 맞게 surface anchor와 query를 분리하고 query에서만 surface loss를 계산하도록 구성한 random 모델이다. curvature 모델은 동일한 모델 구조와 학습·평가 조건을 유지한 상태에서 surface anchor 선택 규칙만 곡률 가중 방식으로 변경하였다. 이에 따라 두 모델의 차이는 surface anchor 선정 방식으로 제한되며, 곡률 정보를 반영한 anchor 배치가 표면 및 체적 유동장 예측에 미치는 영향을 비교하였다.

2.2 국소 형상 변화와 곡률 가중 샘플링

곡률 가중치는 CFD 정답 물리량을 사용하지 않고 표면 좌표만으로 계산한다. 각 표면점에서 자기 자신을 제외한 k 개의 최근접 이웃(k-nearest neighbors, k-NN)을 탐색하고, 이웃 좌표의 주성분 분석(principal component analysis, PCA)에 기반한 surface variation을 구한다. 이러한 공분산 분석은 국소 점 분포가 접평면에서 벗어나는 정도를 나타내는 데 사용된다.⁽³⁾

표면점 i 의 이웃 집합을 $\mathcal{N}_i$, 이웃 좌표의 평균을 $\bar{x}_i$ 라 하면, 국소 공분산 행렬은 식 (1)과 같다. 이웃 수는 $k=32$ 로 설정하였다.

$$C_i = \frac{1}{k} \sum_{j \in \mathcal{N}_i} (x_j - \bar{x}_i)(x_j - \bar{x}_i)^T \quad (1)$$

공분산 행렬의 고유값은 작은 순서대로 $\lambda_{1,i}$, $\lambda_{2,i}$, $\lambda_{3,i}$ 로 정렬하였다. 이때 $\lambda_{1,i}$ 는 해당 표면점 주변의 인접점들이 국소 표면을 대표하는 평면의 수직 방향으로 퍼진 정도를 나타낸다. 주변 표면이 평탄할수록 인접점은 하나의 평면에 가깝게 분포하므로 $\lambda_{1,i}$ 가 작고, 표면의 굽힘이나 형상 변화가 커질수록 그 값이 증가한다. 이를 바탕으로 surface variation s_i 를 식 (2)와 같이 정의하였다. s_i 는 최소 고유값이 전체 고유값의 합에서 차지하는 비율로 계산되는 무차원 지표이며, 국부 표면의 형상 변화 정도를 나타낸다. 본 연구에서는 s_i 를 Surface anchor의 선택 가중치로 사용하였다. ε 은 분모가 0에 가까워지는 경우의 수치적 불안정을 방지하기 위한 작은 양의 상수이다.

$$s_i = \frac{\lambda_{1,i}}{\lambda_{1,i} + \lambda_{2,i} + \lambda_{3,i} + \varepsilon} \quad (2)$$

표면점 i 의 국부 형상 특성은 인접점 좌표의 공분산 행렬로부터 얻은 고유값을 이용하여

surface variation s_i 로 정의하였다. s_i 는 최소 고유값이 전체 고유값의 합에서 차지하는 비율로 계산되는 무차원 지표이며, 국부 표면이 평면에 가까울수록 작은 값을 갖는다.

이하 곡률 기반 sampling은 이러한 surface variation에 따라 sampling weight를 부여하는 방법을 의미한다. 본 연구에서는 인접점 수를 $k = 32$ 로 설정하여 surface variation을 계산하였다.

표면점의 수를 N 이라 할 때, 각 점의 초기 선택 가중치는 식 (3)과 같이 균등 sampling 성분과 곡률 기반 성분의 가중합으로 정의하였다. 여기서 δ 는 곡률 기반 가중치가 0이 되는 것을 방지하기 위한 작은 양의 상수이며, 곡률 정보의 반영 비율을 결정하는 가중계수 α 는 본 실험에서 0.75로 설정하였다. Fig. 1은 동일한 차량 표면에 대해 random sampling과 curvature-weighted sampling으로 선택된 surface anchor의 공간 분포를 비교한 결과로, 곡률 가중치 적용에 따라 형상 변화가 큰 영역에서 anchor의 선택 빈도가 증가하는 경향을 보여준다.

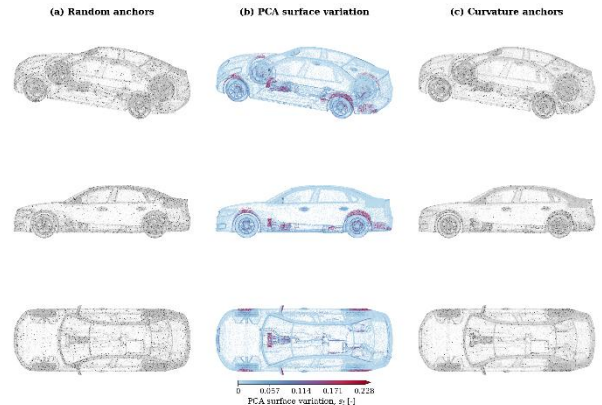


Fig. 1 Comparison of random and curvature-weighted surface anchor distributions based on PCA surface variation

$$w_i = \frac{1-\alpha}{N} + \alpha \frac{s_i + \delta}{\sum_{j=1}^N (s_j + \delta)} \quad (3)$$

표면 전체에서 query point를 먼저 균등하게 추출한 후, query에 포함되지 않은 표면점에서 surface anchor를 선택하였다. 이때 anchor 선택에는 식 (3)의 가중치를 적용하여 곡률이 큰 표면점이 더 높은 확률로 선택되도록 하였다. query point는 전체 표면을 균등하게 대표하도록 유지하고, 곡률 정보는 surface anchor를 선택하는 과정에만 반영하였다.

가중계수 α 는 surface anchor를 선택할 때 곡률 정보를 얼마나 강하게 반영할지를 조절한다. 본 실험에서는 $\alpha = 0.75$ 로 설정하여 곡률이 큰 표면점의 선택 가능성을 높이면서도, 전체 표면에서 anchor가 선택될 수 있도록 균등 sampling 성분을

함께 유지하였다. surface anchor는 한 번 선택된 점이 다시 선택되지 않는 비복원 방식으로 추출하였다.

각 형상에 대한 곡률 지표는 인접점 분포를 이용하여 한 번 계산한 후 저장하여 재사용하였다. 이를 통해 동일한 형상에 대해 국소 공분산과 surface variation을 학습 반복마다 다시 계산하지 않도록 하여, 곡률 기반 sampling에 추가되는 반복 계산을 줄였다.

2.3 데이터 구성과 학습 조건

실험에는 DrivAerML을 10배 서브샘플링하여 구성된 데이터를 사용하였다⁽⁴⁾.

학습 집합은 case 1-50의 50개 형상, 검증 집합은 case 10개 형상으로 구성하였다. 시험 집합은 case 중 자료가 없는 case 3개를 제외한 50개 형상으로 구성하였다. 두 모델에는 동일한 데이터 분할을 적용하였으며, 물리량 정규화에는 학습 집합에서 계산한 동일한 통계값을 사용하였다.

표면 입력에는 16,384개의 surface anchor와 8,192개의 surface query를 사용하였으며, 체적 입력에는 16,384개의 volume anchor를 사용하였다. 학습 단계의 surface query는 전체 표면에서 균등 무작위로 선택하고, 검증 및 시험 단계에서는 각 형상에 대해 미리 저장한 8,192개의 고정 균등 query를 두 모델이 동일하게 사용하였다. 따라서 pressure와 wall shear stress의 예측 성능은 두 모델에서 동일한 표면 위치를 기준으로 비교하였다.

두 모델의 차이는 surface anchor의 선택 방법에 두었다. random 모델은 균등 무작위 sampling을 사용하고, curvature 모델은 곡률 가중 sampling을 사용하였다. 이 선택 규칙은 학습뿐 아니라 검증과 시험에도 동일하게 적용하였다. 따라서 본 실험은 동일한 query 위치에서 평가 조건을 유지하면서, random과 curvature 모델이 학습 및 추론 과정에서 서로 다른 surface anchor 분포를 사용할 때 발생하는 예측 성능의 차이를 비교하도록 구성하였다.

모델의 학습 대상은 surface pressure, wall shear stress, total pressure coefficient 및 velocity의 네 물리량으로 설정하였다. 표면 출력은 pressure 1성분과 wall shear stress 3성분으로 구성하였으며, 체적 출력은 total pressure coefficient 1성분과 velocity 3성분을 사용하였다. 모델의 체적 출력 구조에는 vorticity 3성분이 포함되어 있으나, vorticity는 학습 손실 계산에서 제외하였으며 본 연구의 성능 비교에도 사용하지 않았다. 두 방법에 공통으로 적용한 학습 및 평가 조건은 Table 1에 정리하였다.

Table 1 common experimental settings

Parameter	Value
Train / validation / test	50 / 10 / 50
Epochs / seed / batch size	200 / 42 / 1
Surface anchors / queries	16,384 / 8,192
Volume anchors / queries	16,384 / 0
Geometry points / supernodes	32,768 / 16,384
Hidden dimension / heads	192 / 3
Surface / volume blocks	12 / 12
Optimizer / betas	Lion / 0.9, 0.99
Initial / final learning rate	5×10^{-5} / 1×10^{-6}
Warmup / weight decay	5% / 0.05
Gradient clipping norm	1.0
Training precision	Mixed float16
Evaluation interval	10 epochs

학습 손실은 정규화된 네 물리량의 평균제곱오차(mean squared error, MSE)를 합한 식 (4)로 정의하였다. 여기서 $\mathcal{F} = \{p, \tau, C_p, u\}$ 는 학습 대상 물리량의 집합, n_f 는 손실 계산점 수, d_f 는 성분 수이며, $\hat{y}_f$ 는 정규화된 값을 나타낸다. 스칼라 물리량의 d_f 는 1, 벡터 물리량의 d_f 는 3이다.

$$L = \sum_{f \in \mathcal{F}} \frac{\|\hat{y}_f - y_f\|_2^2}{n_f d_f} \quad (4)$$

표면 손실은 surface anchor와 분리된 surface query에서만 계산하였으며, 체적 손실은 volume anchor에서 계산하였다. 이 구성은 surface anchor의 선택 방식이 표면 손실을 계산하는 정답 위치의 분포까지 직접 변화시키는 것을 방지하고, 두 모델이 동일한 surface query를 기준으로 학습 및 평가되도록 하기 위해 적용하였다. EMA 가중치는 사용하지 않았다. 검증 집합의 surface pressure relative L2 error를 기준으로 최적 체크포인트를 선택한 결과 두 모델 모두 200 epoch의 가중치가 선택되었다. 시험 평가는 해당 epoch의 모델 가중치를 사용하여 수행하였다. 실험은 NVIDIA L4 GPU 환경에서 수행하였으며, 200 epoch 학습에는 random 모델 약 5.01 시간, curvature 모델 약 5.07 시간이 소요되었다.

2.4 평가 지표와 비교 범위

평가는 예측값과 정답값을 동일한 점에 대응시키고 정규화를 역변환한 뒤 수행하였다. 형상 c에서 물리량 f의 예측 및 정답 배열을 각각 $\hat{Y}_{cf}$ 와 Y_{cf}

라 할 때, 시험 형상 수 C 에 대한 평균 MSE와 상대 L2 오차 R_f 는 식 (5), (6)과 같다.

$$MSE_f = \frac{1}{C} \sum_{c=1}^C \frac{\|\hat{Y}_{cf} - Y_{cf}\|_F^2}{n_{cf} d_f} \quad (5)$$

$$R_f = \frac{100}{C} \sum_{c=1}^C \frac{\|\hat{Y}_{cf} - Y_{cf}\|_F}{\|Y_{cf}\|_F} [\%] \quad (6)$$

두 지표는 각 시험 형상에서 개별적으로 계산한 후, 50개 형상의 결과를 산술평균하였다. 벡터 물리량의 Relative L2 error는 모든 평가점에서 세 성분을 함께 포함한 Frobenius norm으로 계산하였으며, 벡터의 크기만을 이용한 오차와 구분하였다. Surface metric은 각 형상에 고정된 8,192개의 균등 Surface query에서 계산하므로, 전체 표면을 적분한 오차나 곡률 크기에 따른 구간별 오차를 의미하지 않는다. 또한 Surface query와 곡률 점수는 모두 원래 표면점에서 계산되므로, 각 위치의 sampling 분포는 원본 표면점의 공간 밀도에 영향을 받는다.

정량 비교에는 50개 테스트 형상의 정규 평가 결과를 사용하였다. 형상별 metric을 다시 산술평균하여 저장된 요약값과 일치함을 확인하였으며, Relative L2 계산에 사용되는 정답 norm도 두 방법에서 동일하게 확인되었다. 다만 모든 시험 형상에 대해 원본 point index를 직접 대조하는 검증은 수행하지 않았다. 또한 일부 저장된 시각화 결과에서는 형상 입력과 Volume sampling 위치가 두 방법 사이에서 일치하지 않았으므로, 해당 시각화는 Random과 Curvature 간 공간적 예측 분포를 직접 비교하는 근거에서 제외하였다.

2.5 예측 결과와 고찰

Table 2는 200 epoch에서 평가한 시험 집합의 relative L2 error를 나타낸다. curvature 모델의 surface pressure 오차는 random 대비 0.01086%p 증가하였으며, wall shear stress 오차는 0.05722%p 증가하였다. 반면 total pressure coefficient와 velocity 오차는 각각 0.14058%p와 0.14903%p 감소하였다. 따라서 surface anchor의 선택 방식을 곡률 가중 방식으로 변경한 본 실험에서는 surface pressure와 wall shear stress의 평균 오차 감소는 확인되지 않았으며, volume 물리량에서는 상대적으로 작은 폭의 오차 감소가 나타났다.

Table 2 Test relative L2 errors at epoch 200

Quantity	Random (%)	Curvature (%)	Δ (%p)
Surface pressure	9.38416	9.39502	+0.01086
Wall shear stress	17.32103	17.37825	+0.05722
Total pressure coefficient	12.94010	12.79951	-0.14058
Volume velocity	12.19369	12.04466	-0.14903

여기서 Δ 는 curvature-random이며 음수는 오차 감소를 뜻한다. 형상별로는 pressure가 24/50개, wall shear stress가 21/50개, total pressure coefficient가 31/50개, velocity가 35/50개에서 감소하였다. 따라서 표면 물리량은 절반 이하의 형상에서만 개선되었고, 체적 물리량은 과반수의 형상에서 개선되었다.

Table 3 Test MSE in physical space at epoch 200

Quantity (unit)	Random	Curvature
Surface pressure (Pa ²)	1.10564×10^3	1.11142×10^3 (+0.52%)
Wall shear stress (Pa ²)	8.876×10^{-2}	8.947×10^{-2} (+0.80%)
Total pressure coefficient (—)	4.72×10^{-3}	4.62×10^{-3} (-2.12%)
Volume velocity (m/s) ²	3.34142×10^0	3.26216×10^0 (-2.37%)

Table 3의 물리 단위 MSE에서는 곡률 기반 sampling 적용 후 surface 오차가 증가하고 volume 오차는 감소하였다. 전체 정규화 test loss는 0.10869에서 0.10795로 소폭 감소하였으나, 이 값은 네 물리량의 정규화 손실을 합산한 결과이므로 surface 예측 성능은 pressure와 wall shear stress의 개별 지표를 기준으로 판단할 필요가 있다. 물리 단위 MSE 역시 물리량마다 단위와 값의 범위가 다르므로, 서로 다른 물리량의 절대값을 직접 비교하기보다 동일 물리량에서 random과 curvature 결과의 변화를 비교하는 지표로 사용하였다.

곡률 기반 sampling은 차량 표면의 국부적인 형상 변화를 기준으로 anchor의 배치를 조정한다. 따라서 모서리나 곡률 변화가 큰 영역에는 더 많은 anchor가 선택되는 반면, 동일한 anchor 수를 유지하는 조건에서는 상대적으로 완만한 표면에 배치되는 anchor 수가 감소한다. 그러나 본 연구에서 사용한 곡률 점수는 표면 좌표의 기하학적 변화만을 반영하며, pressure 또는 wall shear stress의 오차와 공간적 구배를 직접 사용하지 않는다. 실제 유동 변화가 큰 위치와 기하학적 곡률이 큰 위치가 일치하는 정도에 따라 sampling 효과가 달라

질 수 있으며, 이번 결과에서는 곡률 정보를 강화한 anchor 배치가 surface 예측 오차 감소로 이어지지 않았다.

향후 연구에서는 동일한 평가점을 곡률 수준에 따라 구분하여 고곡률 영역과 완만한 영역의 pressure 및 wall shear stress 오차를 각각 비교하고, 영역별 surface anchor 수와 공간 분포를 함께 분석할 필요가 있다. 이를 통해 곡률 가중 sampling이 실제로 형상 변화가 큰 영역의 anchor 배치와 국소 예측 오차에 어떠한 영향을 주는지 확인할 수 있다. 또한 현재 결과는 $\alpha = 0.75$, $k = 32$, random seed 42의 단일 실험 random-curvature 비교에 해당하므로, α 와 k 변화에 따른 민감도 분석, 복수 random seed 반복 실험 및 전체 격자 기반 평가를 추가하여 결과의 재현성과 적용 범위를 검토할 필요가 있다.

3. 결 론

본 연구에서는 AB-UPT의 surface anchor 선정 과정에 PCA 기반 국소 surface variation을 적용하고, 기존의 균등 random sampling과 비교하였다. 두 모델에는 동일한 균등 surface query와 query 전용 손실을 적용하여 anchor의 배치와 표면 정답점의 선택을 분리하였으며, anchor 선정 방식 이외의 모델 구조와 주요 학습·평가 조건은 동일하게 유지하였다.

200 epoch 시험 결과에서 curvature 모델의 surface pressure와 wall shear stress relative error는 random 모델보다 각각 0.11573%와 0.33035% 증가하였으며, total pressure coefficient와 velocity는 각각 1.08647%와 1.22219% 감소하였다. 곡률 정보를 반영한 surface anchor 재배치는 평균 surface 예측 오차의 감소로 이어지지 않았으나, volume 물리량에서는 일관된 방향의 소폭 감소가 나타났다. 이러한 결과는 surface anchor의 위치 변화가 surface 예측에만 국한되지 않고 AB-UPT 내부의 surface-volume 정보 전달에도 영향을 줄 수 있음을 보여준다. 곡률 가중 sampling에서는 형상 변화가 큰 표면점이 더 자주 선택되므로 해당 위치의 기하학적 특징이 surface branch를 통해 모델의 공통 표현에 상대적으로 많이 반영된다. 자동차 외부유동에서 이러한 국부 형상은 차체 주변 유동의 방향 변화와 압력장 형성에 관여하므로 surface에서 강조된 형상 정보가 인접한 volume flow field의 표현에도 영향을 주었을 가능성이 있다. 반면 surface pressure와 wall shear stress는 표면 형상뿐 아니라

국부 압력구배, 경계층 발달 및 박리와 같은 유동 상태에도 좌우된다. 따라서 기하학적 곡률만을 기준으로 anchor를 집중하는 방식은 surface 오차가 큰 위치와 일치하지 않을 수 있으며 이번 실험에서 surface 오차가 감소하지 않은 결과와도 연결된다. volume 오차 감소의 원인을 보다 명확히 확인하려면 surface-volume 간 정보 전달 경로와 anchor 분포의 영향을 분리한 ablation이 추가로 필요하다.

후 기

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No. RS-2024-00355857)

참고문헌

- (1) Ashton, N., Mockett, C., Fuchs, M., Fliessbach, L., Hetmann, H., Knacke, T., Schonwald, N., Skaperdas, V., Fotiadis, G., Walle, A., Hupertz, B. and Maddix, D., 2024, "DrivAerML: High-Fidelity Computational Fluid Dynamics Dataset for Road-Car External Aerodynamics," arXiv:2408.11969.
- (2) Alkin, B., Bleeker, M., Kurle, R., Kronlachner, T., Sonnleitner, R., Dorfer, M. and Brandstetter, J., 2025, "AB-UPT: Scaling Neural CFD Surrogates for High-Fidelity Automotive Aerodynamics Simulations via Anchored-Branched Universal Physics Transformers," Transactions on Machine Learning Research.
- (3) Pauly, M., Gross, M. and Kobbelt, L. P., 2002, "Efficient Simplification of Point-Sampled Surfaces," Proceedings of IEEE Visualization 2002, pp. 163~170.
- (4) Emmi AI, "DrivAerML Subsampled 10x," Hugging Face Datasets, accessed September 7, 2026.