

Integracja danych z wielu sensorów

Metody Sztucznej Inteligencji w Robotyce

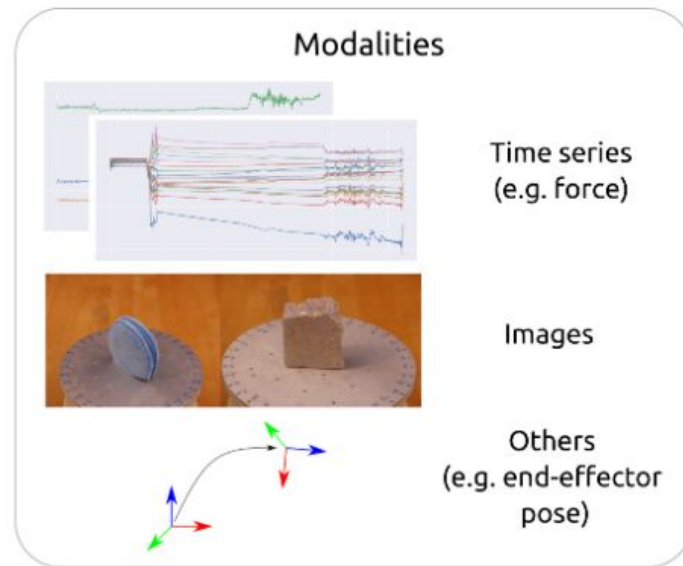
Autorzy: Dariusz Max Adamski, Sławomir Gilewski

Problem

Różne modalności danych

- Obraz
- Dźwięk
- Sensory (pozycja, rotacja, siły)
- Wiedza?

Jak je zintegrować w jednym modelu?

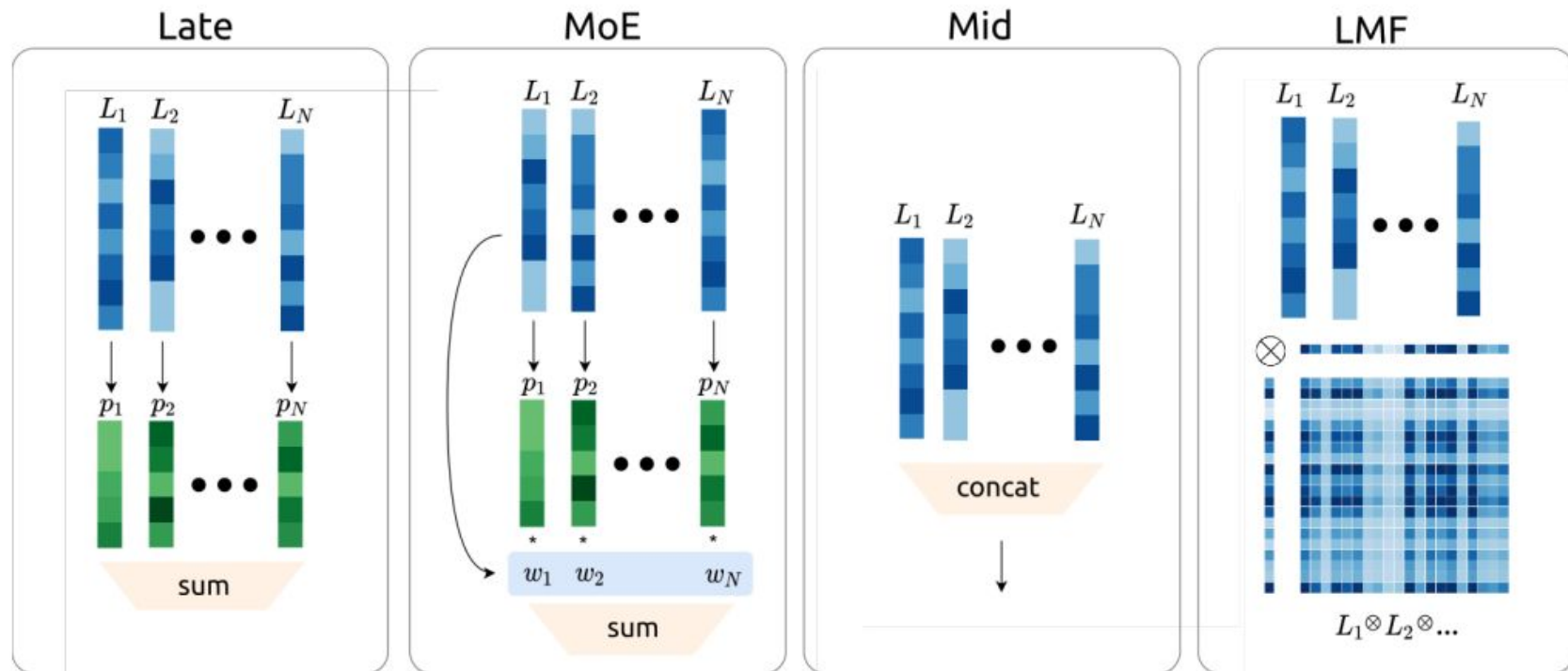


Przegląd klasycznych podejść (Luo)

TABLE II
SUMMARIZATION OF ADVANTAGES AND LIMITATIONS OF FUSION METHODS AND ALGORITHMS

Methods		Advantages	Limitations
Estimation Methods	Kalman Filter	<ul style="list-style-type: none"> • High computational efficiency. • Easy to implement. 	<ul style="list-style-type: none"> • Restricted to linear and Gaussian assumptions. • Lower accuracy and additional phase delays in distributed tracking problems.
	Extended Kalman Filter	<ul style="list-style-type: none"> • Computationally efficient. • Intuitive and easy to use. • Stable in practice estimation problems. 	<ul style="list-style-type: none"> • Non-Gaussian noise processes are not allowed. • System and measurement models need to be differentiable.
	Covariance Intersection	<ul style="list-style-type: none"> • Can fuse independent sensory data with unknown correlation that KF is awkward with. • Can fuse any pair of Gaussian probability density functions. 	<ul style="list-style-type: none"> • Constrained to be linear form of local sensor estimates. • Cannot perform a partial update in estimates.
	Covariance Union	<ul style="list-style-type: none"> • Can avoid the information corruption and eliminate spurious estimations. • Tolerant of inconsistent data sources. 	<ul style="list-style-type: none"> • High computationally demanding. • Inappropriately conservative fusion result may be occurred.
Classification Methods	Support Vector Machine	<ul style="list-style-type: none"> • Can deal with non-linear and non-monotonic data. • Can deliver a unique solution. 	<ul style="list-style-type: none"> • With the lack of transparency of results. • Problems in the choice of the kernel function. • High complexity and extensive memory requirements in large-scale tasks.

State-of-the-art (Bednarek)



Eksperymenty (Bednarek)

BioTac Grasp Stability Dataset (BiGS)

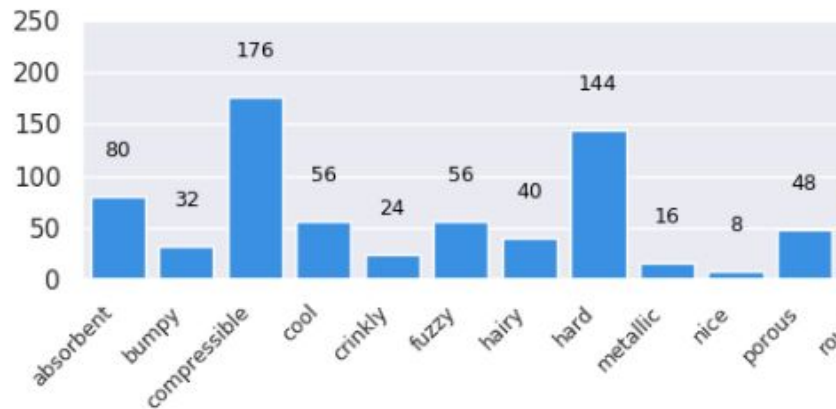
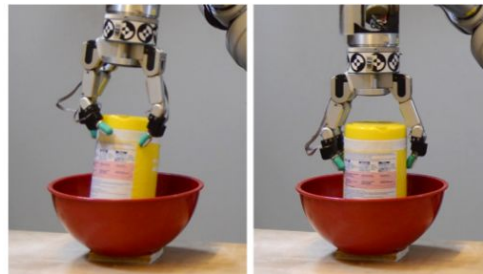
- Predykcja czy robot upuści obiekt
- Pozycja, obrót, siły rączki, czujniki dotyku i momentu obrotowego

Penn Haptic Texture Toolkit (HaTT)

- Rozpoznawanie tekstur
- Obrazki tekstur i szeregi czasowe z danymi haptycznymi (dotyk)

Penn Haptic Adjective Corpus 2 (PHAC-2)

- Wieloetykietowa klasyfikacja własności tekstur
- Zdjęcia obiektów z różnych kątów, szeregi czasowe z danymi haptycznymi (ściskanie)



Eksperymenty (Bednarek)

Dane były dodatkowo zaszumione

- BiGS: Nieznaczne różnice pomiędzy metodami (ACC=88%)
- HaTT: Late Fusion najlepsze (ACC=79.5%), reszta też dobrze (ACC ok. 77%)
- PHAC-2: Late, MoE i Mid najlepsze (AUROC=0.92), LMF (AUROC=0.9)

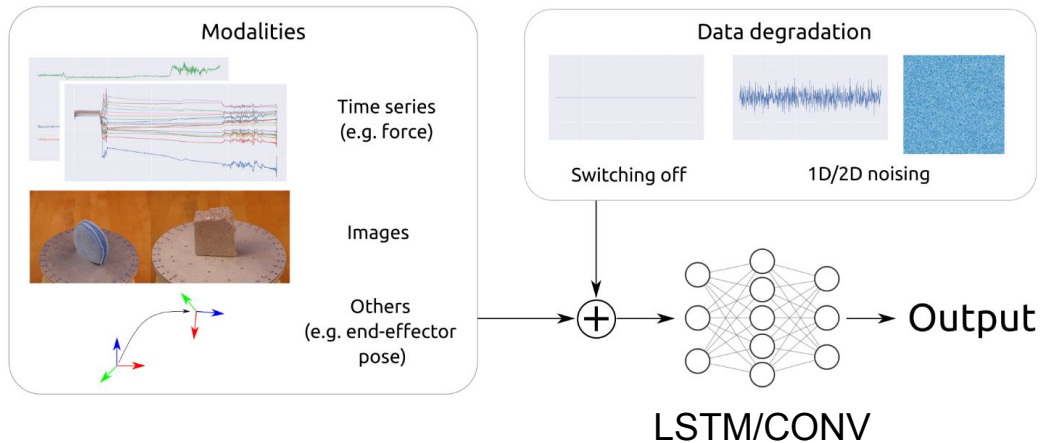


Table 7. Results obtained for models trained on datasets containing noised/zeroed inputs from the leading modality and a zeroed leading modality during tests.

	Late	MoE	Mid	LMF
BiGS [%]	85.77	88.76	86.27	89.01
HaTT [%]	54.85	57.05	53.95	53.25
PHAC-2 [-]	0.23	0.24	0.21	0.61

Przykład architektury end-to-end (Liang)

- Różne sensory (na dole) i zadania (na górze) dopełniają się nawzajem.
- Autorzy proponują model godzący ze sobą dwa sensory oraz cztery zadania.
- Udowodnili, że docelowe zadanie - detekcja obiektów w 3d zyskuje na uczeniu na podstawie kilku podzadani i fuzji danych z kilku sensorów.

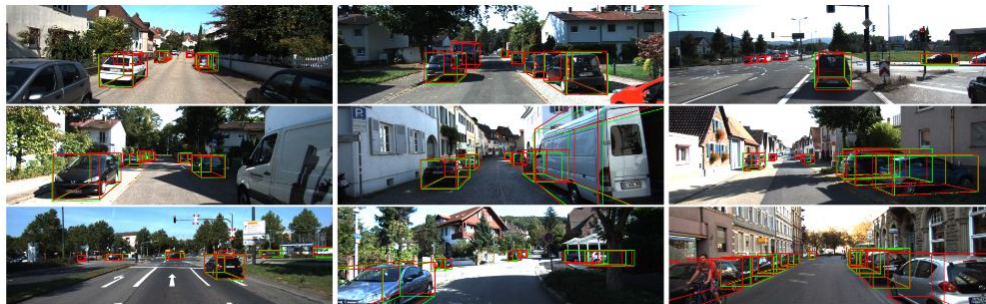
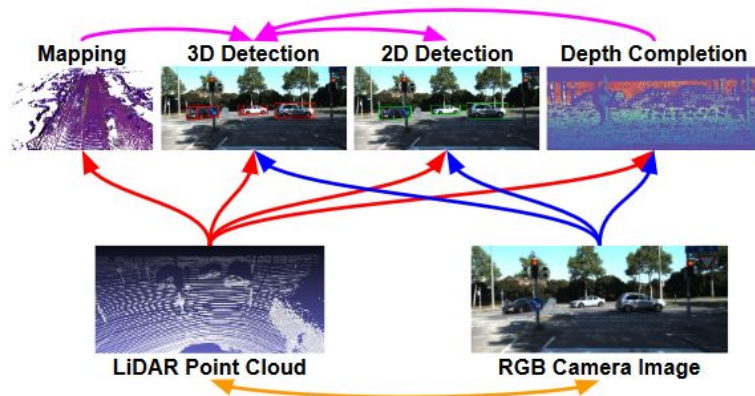
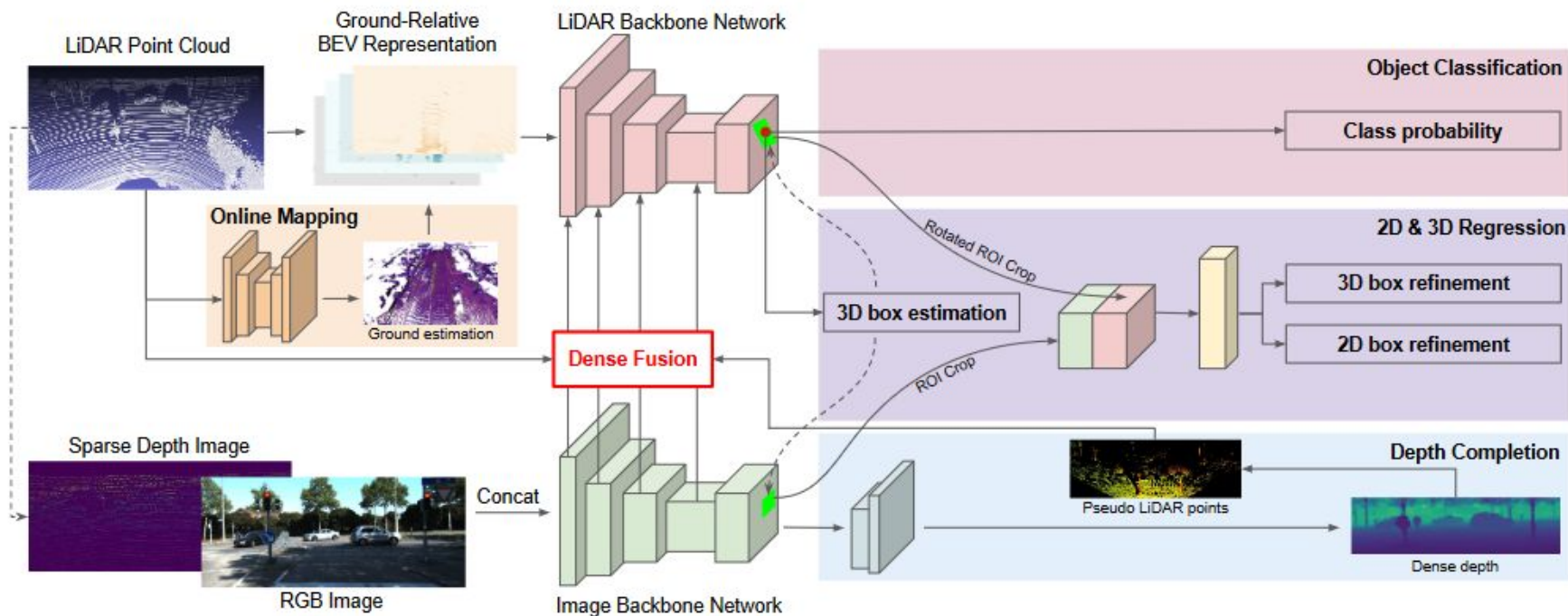


Figure 6. Qualitative results of 3D object detection (car) on KITTI benchmark. We draw object labels in green and our detections in red.




Architektura łącząca sensory i podproblemy (Liang)



Wyniki ich architektury (Liang)

Detector	Input Data		Time (ms)	2D AP (%)			3D AP (%)			BEV AP (%)		
	LiDAR	IMG		easy	mod.	hard	easy	mod.	hard	easy	mod.	hard
SHJU-HW [30, 7]		✓	850	90.81	90.08	79.98	-	-	-	-	-	-
RRC [19]		✓	3600	90.61	90.23	87.44	-	-	-	-	-	-
MV3D [5]	✓		240	89.80	79.76	78.61	66.77	52.73	51.31	85.82	77.00	68.94
VoxelNet [31]	✓		220	-	-	-	77.49	65.11	57.73	89.35	79.26	77.39
SECOND [27]	✓		50	90.40	88.40	80.21	83.13	73.66	66.20	88.07	79.37	77.95
PIXOR [29]	✓		35	-	-	-	-	-	-	87.25	81.92	76.01
PIXOR++ [28]	✓		35	-	-	-	-	-	-	89.38	83.70	77.97
HDNET [28]	✓		50	-	-	-	-	-	-	89.14	86.57	78.32
MV3D [5]	✓	✓	360	90.53	89.17	80.16	71.09	62.35	55.12	86.02	76.90	68.49
AVOD [12]	✓	✓	80	89.73	88.08	80.14	73.59	65.78	58.38	86.80	85.44	77.73
ContFuse [13]	✓	✓	60	-	-	-	82.54	66.22	64.04	88.81	85.83	77.33
F-PointNet [16]	✓	✓	170	90.78	90.00	80.80	81.20	70.39	62.19	88.70	84.00	75.33
AVOD-FPN [12]	✓	✓	100	89.99	87.44	80.05	81.94	71.88	66.38	88.53	83.79	77.90
Our MMF	✓	✓	80	91.82	90.17	88.54	86.81	76.75	68.41	89.49	87.47	79.10

Table 1. Evaluation results on the testing set of KITTI 2D, 3D and BEV object detection benchmark (car). We compare with previously published detectors on the leaderboard ranked by Average Precision (AP) in the moderate setting.

105	UberATG-MMF		94.25 %	97.41 %	89.87 %	0.08 s	GPU @ 2.5 Ghz (Python)	<input type="checkbox"/>
M. Liang*, B. Yang*, Y. Chen, R. Hu and R. Urtasun: Multi-Task Multi-Sensor Fusion for 3D Object Detection . CVPR 2019.								
167	UberATG-MMF		77.43 %	88.40 %	70.22 %	0.08 s	GPU @ 2.5 Ghz (Python)	<input type="checkbox"/>
M. Liang*, B. Yang*, Y. Chen, R. Hu and R. Urtasun: Multi-Task Multi-Sensor Fusion for 3D Object Detection . CVPR 2019.								
139	UberATG-MMF		88.21 %	93.67 %	81.99 %	0.08 s	GPU @ 2.5 Ghz (Python)	<input type="checkbox"/>
M. Liang*, B. Yang*, Y. Chen, R. Hu and R. Urtasun: Multi-Task Multi-Sensor Fusion for 3D Object Detection . CVPR 2019.								

Predykcja najlepszej architektury do fuzji danych z sensorów (Molino-Minero-Re)

- Fuzja danych z sensorów spotkała się z wysokim zainteresowaniem, w wyniku czego powstaje wiele różnych podejść lub “architektur” znacząco różniących się od siebie
- W efekcie trudno jest stwierdzić, które podejście jest lepsze do określonego zbioru sensorów i środowiska
- Autorzy proponują metodę będącą w stanie przewidzieć najlepsze podejście do fuzji danych sensorów. M
- A to pozwolić na odejście od metody “prób i błędów”.

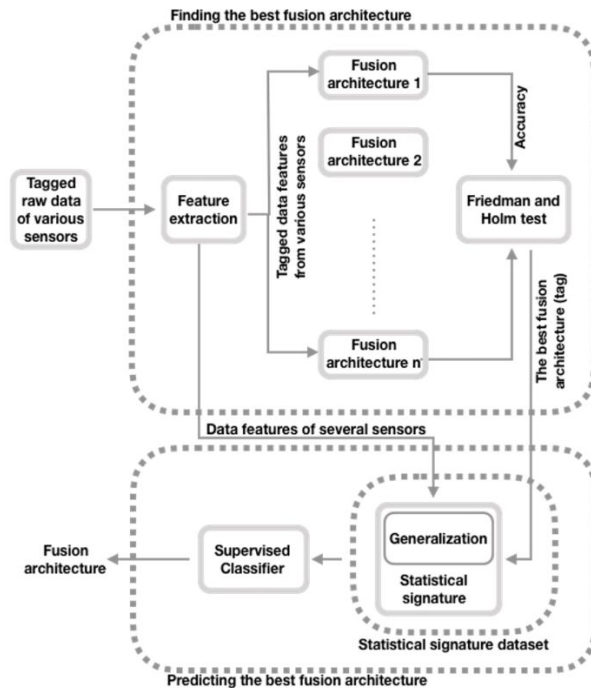


Figure 1. General schematic of the enhanced approach that identifies the optimal fusion architecture (adapted from Brena et al. [12]).

Wyniki predykcji architektury fuzji ()

Table 6. Fusion strategies identification (RFC-based performance results).

Label	Precision		Recall		f1-Score		Support
	Ours	Brena [12]	Ours	Brena [12]	Ours	Brena [12]	
ABWRFC	0.98	1.00	1.00	1.00	0.99	1.00	47
Agg	0.85	0.87	0.85	0.87	0.85	0.87	47
MVSWSF	0.91	0.88	0.91	0.81	0.91	0.84	47
MVSWoSF	0.90	0.90	0.96	0.91	0.93	0.91	47
VotWSF	1.00	0.92	0.91	0.98	0.96	0.95	47
avg/total	0.93	0.91	0.93	0.91	0.93	0.91	235
accuracy					0.93	0.91	235

Bibliografia

1. Luo R. C., Multisensor Fusion and Integration: A Review on Approaches and Its Applications in Mechatronics
2. Bednarek et al., On Robustness of Multi-Modal Fusion—Robotics Perspective
3. Liang et al., Multi-Task Multi-Sensor Fusion for 3D Object Detection
4. Molino-Minero-Re et al., Improved Accuracy in Predicting the Best Sensor Fusion Architecture for Multiple Domains

Dziękujemy za uwagę

