Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.28509

FDConv-Enhanced Multi-Information Fusion for Real-Time GTAW Weld Quality Monitoring

Sheng Guan

School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China
15951892205@163.com

Abstract. Reliable online monitoring of Gas Tungsten Arc Welding (GTAW) is difficult because visual, electrical, and acoustic observations each describe only part of the welding process. In this work, we propose a compact multimodal fusion network with frequency-dynamic convolution (FDConv) to improve weld-state recognition under real-time constraints. The network applies modality-specific spectral enhancement before intermediate fusion, enabling effective integration of synchronized arc current/voltage signals, acoustic emission (AE) spectrograms, and infrared (IR) weld-pool images. On a balanced GTAW dataset, the proposed method achieves an F1-score improvement of about 4–5 percentage points over a tuned CNN–LSTM fusion baseline, while maintaining an added latency of no more than 100 ms at a 10 Hz decision rate. The experimental results show that emphasizing informative frequency components prior to fusion helps retain defect-sensitive patterns and yields more stable recognition performance. These findings support the use of frequency-aware multimodal learning for real-time GTAW quality monitoring.

Keywords: GTAW, weld quality monitoring, multi-sensor fusion, frequency-dynamic convolution, acoustic emission, infrared weld-pool imaging, real-time diagnostics

1. Introduction

In automated Gas Tungsten Arc Welding (GTAW), maintaining stable penetration while avoiding defects such as lack of fusion, undercut, and burn-through is still difficult, especially when process conditions change rapidly. In many cases, signals from a single sensor are not sufficient to capture weak or short-duration abnormal patterns. Previous studies have shown that weld-pool images, arc electrical signals, and acoustic emission (AE) or sound signals can all provide useful information for weld-state assessment, and combining these sources usually leads to more reliable recognition results than using only one modality [1-5]. However, two practical issues remain. First, conventional spatial convolution is not well suited to representing frequency-sensitive patterns, which are particularly important in AE signals and arc fluctuation analysis. Second, when noise, interference, or partial information loss occurs, the contribution of different modalities is not always effectively adjusted during fusion. To address these issues, we introduce FDConv, which divides the feature representation into several spectral sub-bands under a fixed parameter budget and learns modality-specific frequency emphasis before multimodal fusion.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.28509

2. Related work

Previous studies have investigated different sensing routes for online weld monitoring. For vision-based methods, deep transfer learning and attention mechanisms have been used to estimate penetration and weld width from top-side and infrared images [1-3]. For acoustic signals, AE time–frequency representations combined with CNN or LSTM models have shown good performance in identifying penetration conditions and abnormal process behavior [4,5]. More broadly, existing reviews suggest that synchronized image, electrical, and sound signals complement one another, and their fusion at the feature level is generally more reliable than using any single modality alone [1,6-8]. Prior work has also shown that frequency-related processing becomes especially useful when defect cues appear as short-term spectral variations [4,5,9]. This provides the basis for our design, in which frequency-selective convolution is introduced before multimodal feature fusion.

3. Methods

3.1. Sensing and preprocessing

We synchronize four streams: arc current (20 kHz), arc voltage (20 kHz), AE microphone (48 kHz), and IR pool imaging (60 FPS, 640×480). Electrical channels are detrended and z-scored in 0.5 s windows. AE is transformed to log-magnitude STFT tiles (1024-point FFT, 75% overlap) and standardized per frequency bin. IR frames are cropped to a weld ROI and normalized per sequence. Decision windows are produced at 10 Hz with 0.1 s stride, yielding aligned quadruplets $(x^I, x^V, x^A, x^{\mathrm{IR}})$.

Table 1. Sensor suite and primary roles

Sensor	Modality	Rate	Role
Arc current	Electrical	20 kHz	Heat input surrogate; penetration/arc stability cues [1,2].
Arc voltage	Electrical	20 kHz	Arc length/energy fluctuation; instability indicators [1,9].
Acoustic emission	Acoustic	48 kHz	Spatter/droplet events; band-limited defect signatures [4,5,9].
IR pool camera	Thermal/visual	60 FPS	Pool geometry/cooling textures; penetration proxies [2,3].

3.2. FDConv encoders and fusion head

Each modality uses a shallow encoder beginning with FDConv followed by light residual units. FDConv parameterizes a small set of frequency-gated kernels and sums their spatial responses:

$$\mathrm{y} = \sum_{k=1}^{\mathrm{K}} \mathcal{F}^{-1}\left(\mathrm{M_k} \odot \mathcal{F}(\mathrm{w})\right) * \mathrm{x},$$

where $\mathcal{F}$ is the Fourier transform, M_k masks disjoint sub-bands, and w are kernel weights. Encoders output fixed-length embeddings projected to a common width and fused by element-wise summation:

$$\mathrm{F} = \phi\left(\sum_{m\in\{I,V,A,\mathrm{IR}\}} \mathrm{W_m f^m} + \mathrm{b_m}\right),$$

with GELU ϕ. A two-layer head jointly predicts (i) weld state (normal vs. defect family) and (ii) a geometric proxy (e.g., bead width). Training uses cross-entropy (1.0) + Huber (0.5), AdamW, and cosine LR decay.

DOI: 10.54254/2755-2721/2025.28509

3.3. Baselines and constraints

The main baseline is a tuned CNN–LSTM fusion (per-modality CNN encoders feeding an LSTM across windows, with feature concatenation). Single-modality ablations are also reported. We profile latency on an edge-class GPU and target ≤100 ms added latency per decision.

4. Results and discussion

4.1. Overall performance

On a balanced GTAW dataset with realistic noise/disturbance injections, FDConv-fusion improves accuracy and F1 over both the CNN–LSTM baseline and the best single stream.

Table 2. Test-set performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Single-modality (best)	82.4	84.1	78.9	81.3
CNN–LSTM (fusion)	88.7	86.5	90.2	88.3
FDConv–Fusion	**93.2**	**92.1**	**94.3**	**93.2**

Gains are concentrated in recall (fewer missed subtle defects) without sacrificing precision, consistent with the premise that frequency-localized AE/electrical cues and mid–high-frequency IR textures complement one another [9-12].

4.2. Fusion diagnostics—placed mid-results for readability

Figure 1 presents the correlation pattern of the fused features on the validation set. In the heatmap, relatively strong correlations can be observed within the AE branch and within the IR branch, indicating that each modality still preserves its own internal structure after encoding. For the AE stream, the stronger responses are mainly concentrated in several adjacent feature groups derived from the time–frequency representation, while for the IR stream they appear more clearly around features associated with weld-pool boundary variation and local thermal texture. In addition to these within-modality blocks, several off-diagonal regions also show noticeable cross-modal interactions. These regions link AE-related components with slowly varying electrical responses and with texture cues extracted from the IR images. Such a pattern suggests that the fused representation does not simply stack different sensing streams together, but retains both modality-specific information and shared responses across sensors. In practical monitoring, this is useful because one channel may temporarily become unreliable due to arc fluctuation, local overexposure, or sensor disturbance. Under these conditions, the remaining modalities can still provide overlapping evidence for the final prediction, which is in line with earlier observations on multimodal robustness in welding monitoring [13,14].

DOI: 10.54254/2755-2721/2025.28509

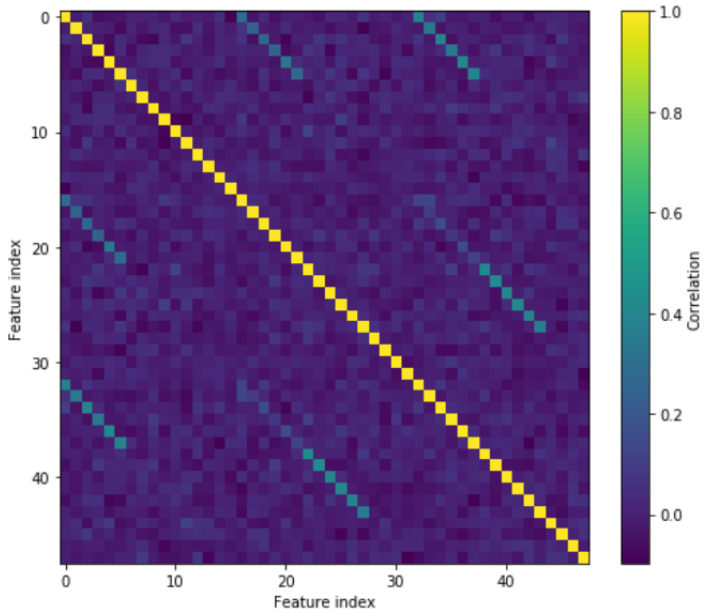


Figure 1. Correlation heatmap of fused features

4.3. Error distributions and robustness—placed later for separation

Figure 2 shows the class-wise absolute error distributions of the baseline model and FDConv, using the bead-width-related estimate as a representative example. The error is calculated as the absolute difference between the predicted value and the reference value obtained from synchronized image-based measurement. Compared with the baseline, FDConv produces a visibly tighter distribution in both the normal class and the defect classes. The interquartile range becomes smaller, the median remains closer to the reference level, and the number of large-error samples is also reduced. This means that the advantage of FDConv is not limited to improving the average result, but also appears in the overall distribution of prediction errors. For online monitoring, this is important because occasional large deviations are often more harmful than a small change in mean performance, especially when the output is used for alarm triggering or process adjustment. The more concentrated distribution therefore suggests that the proposed model is less affected by nuisance factors such as arc-length fluctuation, sensor noise, and short-term disturbance, and it also yields fewer estimates near the decision boundary. Similar improvements in stability and sensitivity have been reported in previous studies that combine visual, electrical, and AE information for weld monitoring [14,15].

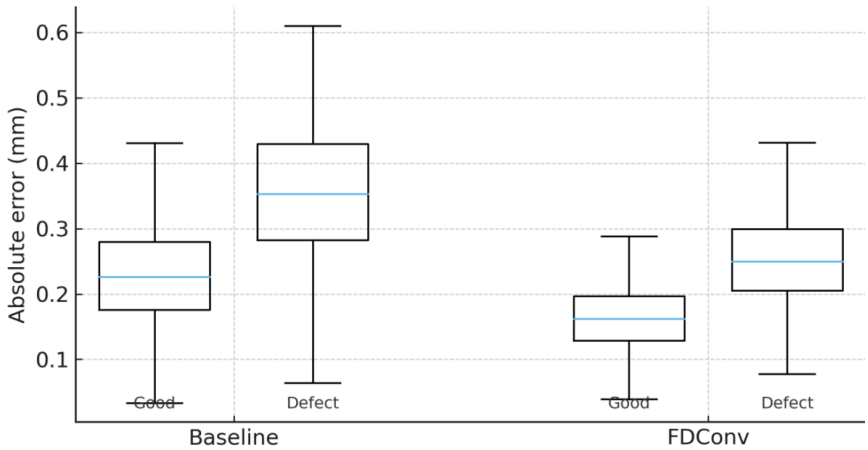


Figure 2. Error distributions by model and class

DOI: 10.54254/2755-2721/2025.28509

4.4. Ablations, throughput, and deployment notes

The ablation study further clarifies the source of the performance gain. When FDConv is replaced with standard convolution, the overall accuracy decreases to about 89%, even though the remaining encoder and fusion settings are kept unchanged. This result indicates that the improvement is closely related to the frequency-aware gating mechanism rather than to the multimodal framework alone. The modality-removal experiments show a similar trend. Removing the AE branch reduces the F1 score by about 2.1 percentage points, while removing the IR branch leads to a drop of about 2.8 percentage points. When both branches are removed, the performance becomes much closer to that of the single-modality configuration, which further confirms that these sensing streams provide complementary information during recognition. From the deployment perspective, latency is measured from the arrival of a complete decision window to the generation of the final prediction. With mixed-precision inference and window-level batching, the added latency remains within 100 ms at a 10 Hz decision rate. In this example setting, the batch size is 8, the decision window length is 0.5 s, and the IR input resolution is 224 × 224. These results indicate that the proposed model remains compatible with the timing requirements of online monitoring and is comparable to previously reported real-time welding vision and multimodal sensing systems.

5. Conclusion

This work presents an FDConv-based multimodal fusion network for GTAW monitoring and demonstrates its effectiveness under real-time constraints. Compared with a strong CNN–LSTM fusion baseline, the proposed method achieves higher recognition accuracy and more concentrated error distributions. By introducing frequency-aware extraction into each modality encoder, the network preserves defect-related spectral information more effectively and combines heterogeneous sensing cues into a compact fused representation. The current study is still limited to the tested materials, joint configurations, and sensing setup. Future research will focus on adaptive sensor weighting, transfer to new welding conditions, and defect localization methods that can provide more direct support for process control.

References

[1] Saimon, A. I., Yangue, E., Yue, X., Kong, Z., & Liu, C. (2025). Advancing additive manufacturing through deep learning: A comprehensive review of current progress and future challenges. IISE Transactions, 1-24.
[2] Xue, B., Du, D., Peng, G., Zhang, Y., Li, R., & Li, Z. (2025). Visual monitoring of weld penetration in aluminum alloy GTAW based on deep transfer learning enhanced by task-specific pre-training and semi-supervised learning. Journal of Manufacturing Processes, 133, 1038-1050.
[3] Nam, K., & Ki, H. (2024). Deep learning-based monitoring system for predicting top and bottom bead widths during the laser welding of aluminum alloy. Journal of Manufacturing Processes, 120, 616-627.
[4] Xia, Y. J., Song, Q., Yi, B., Lyu, T., Sun, Z., & Li, Y. (2025). Improving out-of-distribution generalization for online weld expulsion inspection using physics-informed neural networks. Welding in the World, 69(5), 1309-1322.
[5] Luo, Z., Wu, D., Zhang, P., Ye, X., Shi, H., Cai, X., & Tian, Y. (2023). Laser welding penetration monitoring based on time-frequency characterization of acoustic emission and CNN-LSTM hybrid network. Materials, 16(4), 1614.
[6] Wang, Y., Won, C., & Yoon, J. (2025). A Deep Learning-Based Machine Vision System for Online Monitoring and Quality Evaluation During Multi-Layer Multi-Pass Welding. Sensors (Basel, Switzerland), 25(16), 4997.
[7] Jorge, V. L., Bendaoud, I., Soulié, F., & Bordreuil, C. (2024). Rear weld pool thermal monitoring in GTAW process using a developed two-colour pyrometer. Metals, 14(8), 937.
[8] Wang, Z., Shi, Y., Cui, Y., & Yan, W. (2024). Three-Dimensional Weld Pool Monitoring and Penetration State Recognition for Variable-Gap Keyhole Tungsten Inert Gas Welding Based on Stereo Vision. Sensors (Basel, Switzerland), 24(23), 7591.

DOI: 10.54254/2755-2721/2025.28509

[9] Wolf, C., Sommer, N., & Böhm, S. (2024). Use of a structure-borne sound-based in-process sensor system to identify Weld seam irregularities during electron beam welding. Scientific Reports, 14(1), 22120.
[10] Choi, J., Mazumder, J., & Rice, A. (2020). Innovative additive manufacturing process for successful production of 7000 series aluminum alloy components using smart optical monitoring system (No. 2020-01-1300). SAE Technical Paper.
[11] Shang, F., Cai, D., Sun, B., Han, Y., Yang, M., & Zhan, C. (2025). Research on weld seam feature extraction and defect identification technology. Nondestructive Testing and Evaluation, 40(8), 3339-3376.
[12] Xu, J., Liu, Q., Xu, Y., Xiao, R., Hou, Z., & Chen, S. (2024). Review on the application of the attention mechanism in sensing information processing for dynamic welding processes. Journal of Manufacturing and Materials Processing, 8(1), 22.
[13] Chandrasekhar, N., Muthukumaran, V., & Das, C. R. (2025). Real-time determination of weld penetration status during A-TIG welding of stainless steel employing deep learning approach. Welding in the World, 1-18.
[14] Darwish, A., Persson, M., Ericson, S., Ghasemi, R., & Salomonsson, K. (2025). Weld Defect Detection in Laser Beam Welding Using Multispectral Emission Sensor Features and Machine Learning. Sensors, 25(16), 5120.
[15] Griffin, J. M., Jones, S., Perumal, B., & Perrin, C. (2023, July). Investigating the detection capability of acoustic emission monitoring to identify imperfections produced by the metal active gas (mag) welding process. In Acoustics (Vol. 5, No. 3, pp. 714-745). MDPI.