

Università degli studi di Torino

Dipartimento di informatica



Progetto di Elaborazione di Immagini e Visione Artificiale

Confronto tra Reti Convoluzionali Standard LeNet5 e ConvKAN tramite Grad-CAM

Autore:

Leonardo Magliolo

Anno Accademico 2024/2025

1. Introduzione

Contesto e Obiettivo

In questo progetto, si analizza il comportamento di due modelli di rete convoluzionale, utilizzando il dataset E-MNIST. Questo dataset, derivato da MNIST, include sia cifre sia lettere (maiuscole e minuscole) per un totale di 62 classi, rendendolo un problema di classificazione più complesso rispetto al classico MNIST.

L'obiettivo è confrontare la classica rete convoluzionale LeNet5 con una versione convoluzionale delle Kolmogorov-Arnold Networks (KAN), note come ConvKAN, in termini di interpretabilità delle decisioni. L'utilizzo di Grad-CAM consentirà di visualizzare quali parti dell'immagine hanno maggiormente influenzato la decisione delle due reti.

Introduzione alle KAN

Le Kolmogorov-Arnold Networks (KAN) si basano sul teorema di Kolmogorov-Arnold, che afferma che ogni funzione multivariata continua può essere rappresentata come somma di funzioni univariate. Le KAN sfruttano questo concetto applicando funzioni spline univariate ai pesi, invece che attivazioni fisse sui nodi. I pesi non sono più semplici scalari, ma funzioni spline adattive, in grado di modellare relazioni non lineari più complesse rispetto alle classiche Multi-Layer Perceptrons (MLP). Questo approccio fornisce una rappresentazione più potente e interpretabile, poiché le spline sono facilmente visualizzabili.

Estensione alle ConvKAN

Le ConvKAN estendono i principi delle KAN alle convoluzioni. Invece di filtri lineari rigidi, i filtri sono funzioni spline univariate apprese durante l'addestramento. Parametri come `grid_size` e `spline_order` controllano la complessità di queste spline. Ciò consente alle ConvKAN di catturare pattern più complessi e di fornire, potenzialmente, una miglior interpretabilità.

Motivazione

L'interpretabilità nei modelli di deep learning è cruciale. Grad-CAM aiuta a visualizzare le regioni dell'immagine ritenute importanti dal modello. Confrontare LeNet5 (architettura classica) con ConvKAN (architettura innovativa con spline) è rilevante per capire se la maggiore potenza espressiva dei filtri spline si traduce non solo in potenziali miglioramenti prestazionali, ma anche in una diversa distribuzione dell'attenzione interna.

2. Architetture Utilizzate

LeNet5

LeNet5 segue la sequenza:

- **Conv(1 in, 6 out, 5x5, padding=2) + ReLU**
- **Pool(2x2)** -> da 28x28 a 14x14
- **Conv(6 in, 16 out, 5x5, no padding) + ReLU**
- **Pool(2x2)** -> da 10x10 a 5x5
- **FC(256->120->84->62)** con ReLU intermedie

I filtri sono lineari. Questa rete funge da baseline consolidata.

ConvKAN

La ConvKAN utilizza layer `KAN_Convolutional_Layer` con spline. Il codice riportato definisce la topologia KAN utilizzata nella variante `LeNet5_KAN` (ConvKAN):

- **Primo strato conv KAN:**
 - `in_channels=1, out_channels=6, kernel 5x5, stride=1, padding=0`
 - `grid_size=5, spline_order=3`, attivazione base ReLU
 - **Pool(2x2)** che riduce da 28x28 a 12x12 (dopo conv 5x5 senza padding si ha 24x24, poi pooling dimezza)
- **Secondo strato conv KAN:**
 - `in_channels=6, out_channels=16, kernel 5x5, stride=1, no padding`
 - Stesse impostazioni spline (`grid_size=5, spline_order=3`)
 - **Pool(2x2)** riduce da 8x8 a 4x4
- **Strati Fully Connected:**
 - **Input 256 neuroni** (16 canali da 4x4)
 - **FC(256->120->84->62)** con ReLU intermedie

In questa topologia KAN (ConvKAN), ogni filtro è una funzione spline appresa, rendendo i filtri non lineari e adattivi.

Comparabilità delle Topologie

Le due topologie, LeNet5 e ConvKAN, sono comparabili in termini di struttura generale, ma presentano differenze nei dettagli architetturali:

- **Struttura Generale:**
 - **Entrambe le reti** seguono una struttura simile composta da:
 - **Strati Convolutionali:** Utilizzati per estrarre le caratteristiche dalle immagini.
 - **Strati di Pooling:** Riduzione della dimensionalità spaziale delle feature maps.
 - **Strati Fully Connected:** Per la classificazione finale basata sulle caratteristiche estratte.

Questa similitudine permette un confronto diretto tra le due architetture, poiché operano su livelli simili di elaborazione delle informazioni.

- **Tipo di Filtri Convolutionali:**
 - **LeNet5:**
 - Utilizza **filtri convolutionali lineari** con pesi fissi.
 - I filtri sono semplici mappature lineari che apprendono pesi scalari durante l'addestramento.
 - La linearità dei filtri limita la capacità di catturare pattern complessi non lineari nelle immagini.
 - **ConvKAN:**
 - Integra **filtri convolutionali basati su spline univariate**.
 - I filtri non sono più semplici pesi scalari, ma funzioni spline adattive che possono modellare relazioni non lineari.
 - Questa flessibilità consente ai filtri di adattarsi meglio a pattern complessi e irregolari nelle immagini.
- **Parametrizzazione dei Pesi:**
 - **LeNet5:**
 - I pesi dei filtri sono parametri scalari appresi tramite backpropagation.
 - La struttura dei pesi è rigida, limitando la capacità del modello di adattarsi a variazioni complesse nei dati.
 - **ConvKAN:**

- I pesi sono rappresentati da **funzioni spline**, definite da parametri come `grid_size` e `spline_order`.
 - Le spline permettono una rappresentazione più flessibile e continua dei pesi, migliorando la capacità del modello di catturare variazioni non lineari.
- **Numero di Neuroni nelle Fully Connected Layers:**
 - **LeNet5:**
 - Input alla prima layer fully connected: 256 neuroni.
 - **ConvKAN:**
 - Input alla prima layer fully connected: 256 neuroni .

Entrambe le reti hanno 256 neuroni nelle fully connected layers, garantendo una comparabilità equa nelle fasi finali di classificazione.

Implicazioni delle Differenze

Le differenze architetturali tra LeNet5 e ConvKAN comportano implicazioni significative sulle loro performance e interpretabilità:

- **Performance:**
 - **ConvKAN** può potenzialmente superare LeNet5 in termini di accuratezza su dataset più complessi grazie alla sua maggiore capacità di catturare pattern non lineari.
 - Tuttavia, questo vantaggio può essere attenuato dalla maggiore complessità computazionale e dal rischio di overfitting se non adeguatamente regolarizzato.
- **Interpretabilità:**
 - **LeNet5** offre una maggiore semplicità e stabilità nelle mappe di attivazione, facilitando l'interpretazione dei risultati.
 - **ConvKAN**, pur offrendo una maggiore flessibilità, può generare mappe di attivazione più complesse e meno focalizzate, richiedendo un'analisi più approfondita per interpretare correttamente le decisioni del modello.

In sintesi, mentre LeNet5 e ConvKAN sono comparabili per la loro struttura generale, le differenze nei filtri convoluzionali e nella parametrizzazione dei pesi conferiscono a ConvKAN una maggiore flessibilità e potenzialmente migliori performance su compiti complessi. Tuttavia, questa maggiore flessibilità comporta anche una maggiore complessità computazionale e una interpretabilità più sfaccettata. La scelta tra LeNet5 e ConvKAN dipenderà quindi dalle specifiche esigenze del compito di classificazione e dalle risorse computazionali disponibili.

3. Dataset e Iperparametri

Dataset E-MNIST

- **62 classi** (cifre + lettere)
- **Immagini** 28x28 in scala di grigi
- **Divisione**: train, validation (10% estratto dal train) e test
- **Normalizzazione**: mean=0.1307, std=0.3081

Iperparametri

Gli iperparametri utilizzati sono frutto di ricerche effettuate precedentemente su reti LeNet300 e il loro equivalente KAN. Per un confronto equo:

- **Learning rate**: 0.01
- **Ottimizzatore**: SGD senza momentum
- **Batch size**: 100
- **Epoche**: 50
- **Funzione di perdita**: Cross Entropy
- **Regolarizzazione**: L2 o None
- **Per ConvKAN**: grid_size=5, spline_order=3 costanti

4. Grad-CAM: Implementazione e Adattamenti

Descrizione Grad-CAM

Grad-CAM produce mappe di attivazione pesando le feature map di un layer convoluzionale con i gradienti della classe target. Mostra visivamente dove il modello focalizza l'attenzione.

Adattamenti per ConvKAN

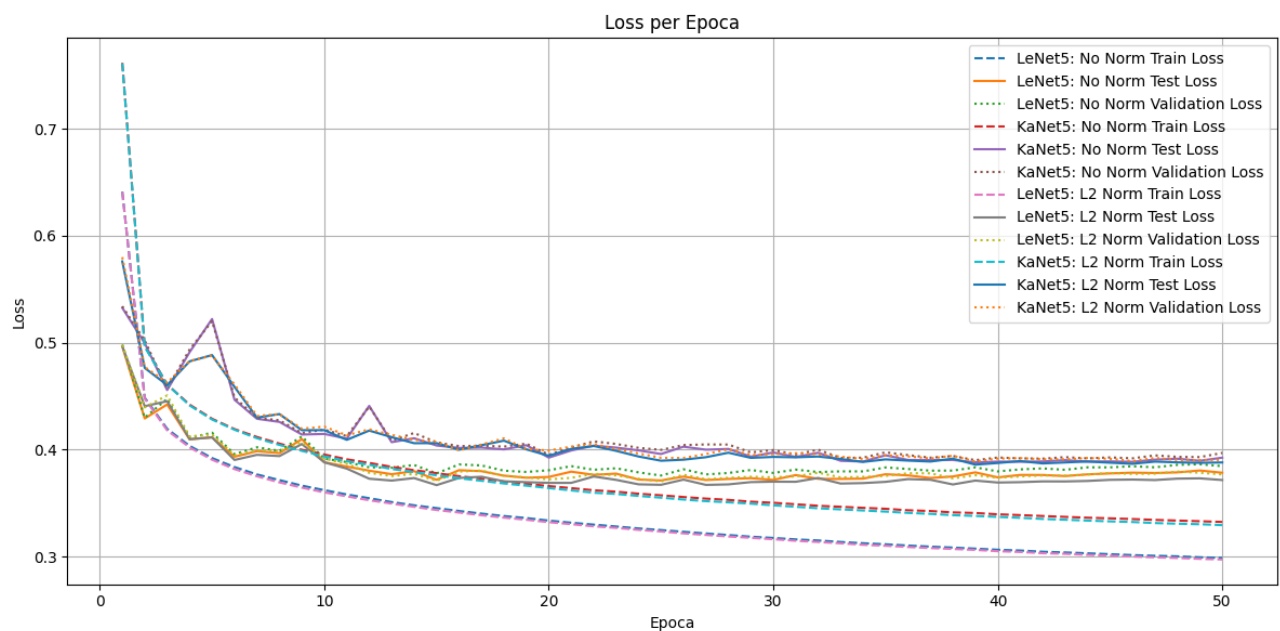
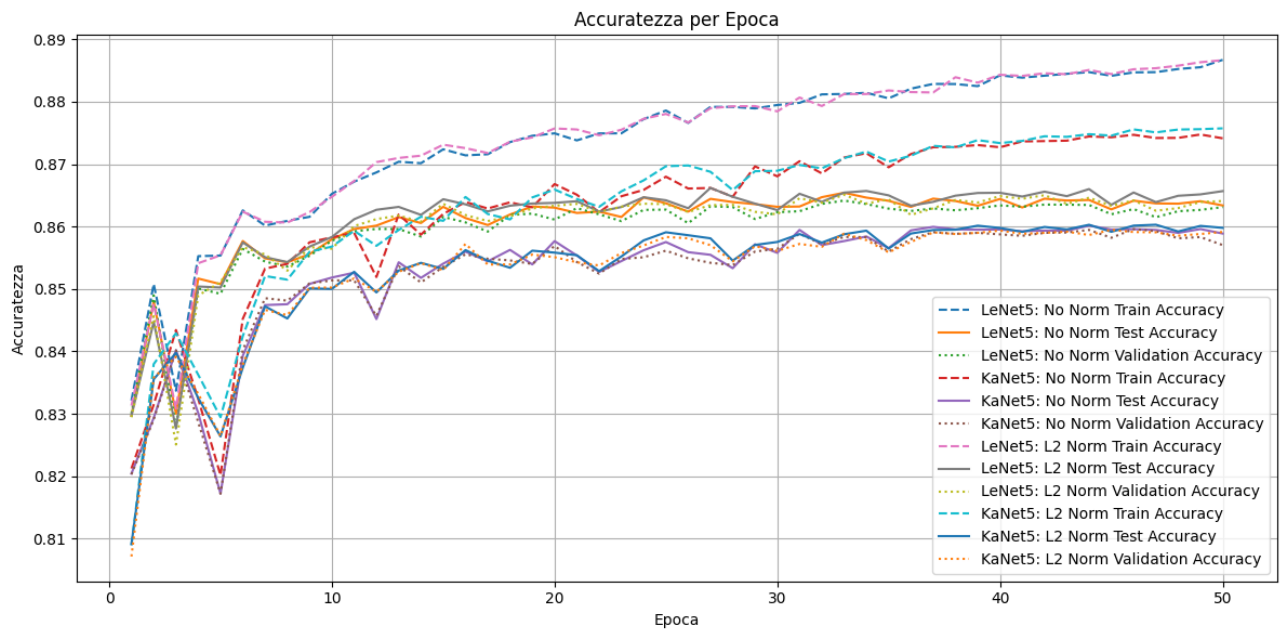
Per ConvKAN, è stato necessario:

- **Identificare il layer conv finale per Grad-CAM**
- **Assicurare che gli hook catturassero correttamente feature map e gradienti dei filtri spline**

5. Risultati

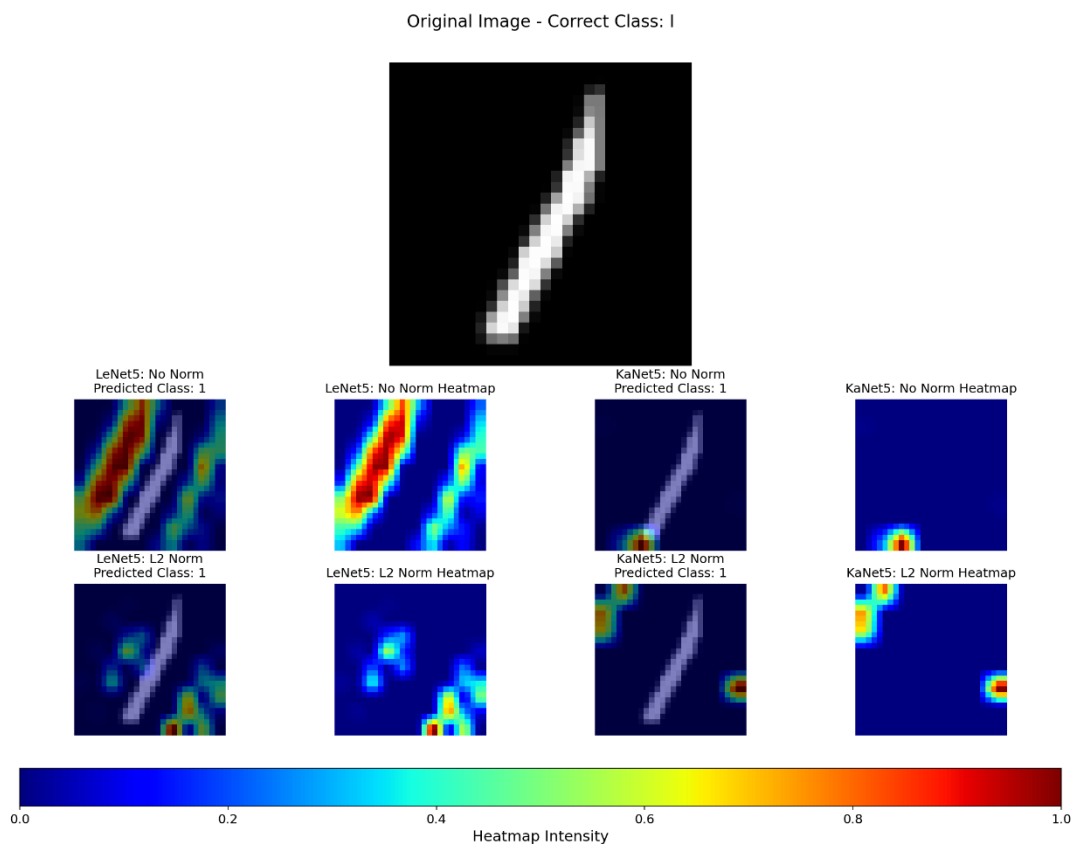
Performance delle Reti

LeNet5 e ConvKAN raggiungono ~90-93% su E-MNIST. LeNet5 “No Norm” ~91%. ConvKAN con L2 può migliorare leggermente. Tuttavia, l’incremento è marginale e richiede maggiore complessità computazionale.

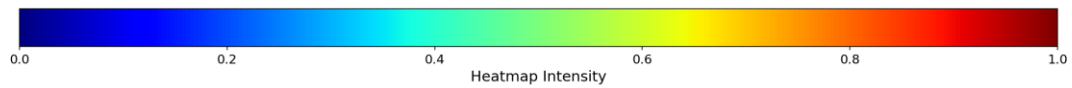
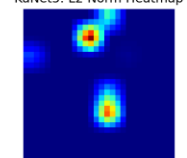
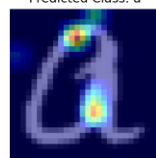
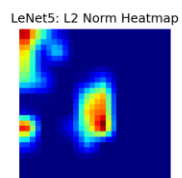
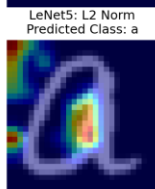
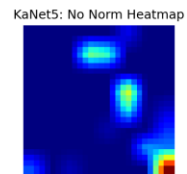
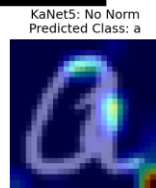
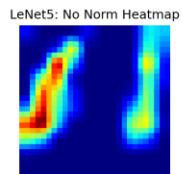
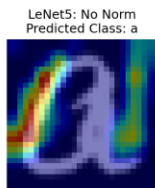
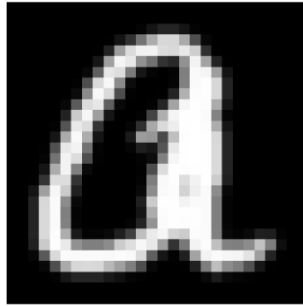


Visualizzazioni delle Heatmap (Grad-CAM)

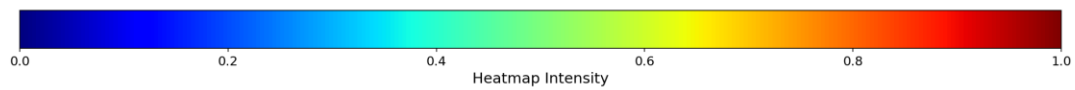
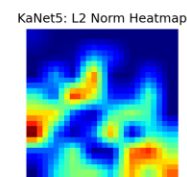
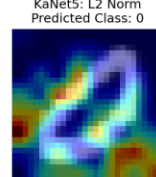
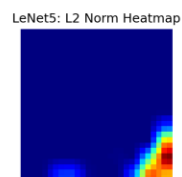
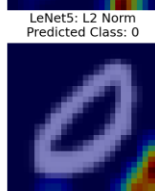
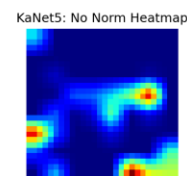
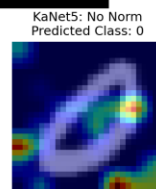
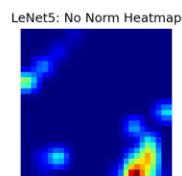
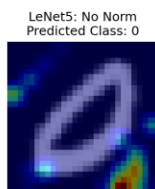
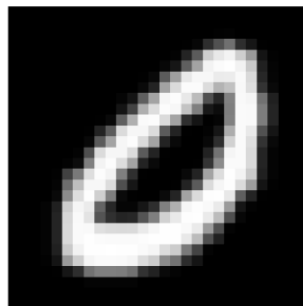
- **Esempi corretti da entrambe:** LeNet5 mostra heatmap concentrate su tratti fondamentali del carattere. ConvKAN focalizzata su tratti principali, ma con attivazioni più diffuse.
- **Esempi errati:** LeNet5 meno nitida ma coerente. ConvKAN più “sparsa”, riflettendo la complessità interna.
- **Esempi in cui una rete sbaglia e l'altra no:** Differenze evidenti nella distribuzione dell'attenzione. ConvKAN può cogliere pattern che LeNet5 ignora (o viceversa).



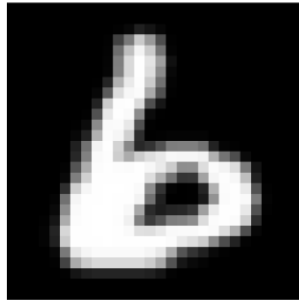
Original Image - Correct Class: a



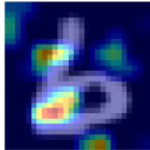
Original Image - Correct Class: 0



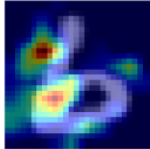
Original Image - Correct Class: 6



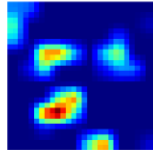
LeNet5: No Norm
Predicted Class: 6



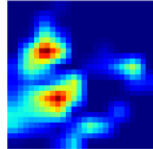
LeNet5: L2 Norm
Predicted Class: 6



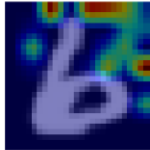
LeNet5: No Norm Heatmap



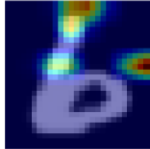
LeNet5: L2 Norm Heatmap



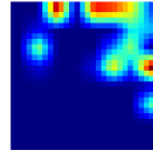
KaNet5: No Norm
Predicted Class: b



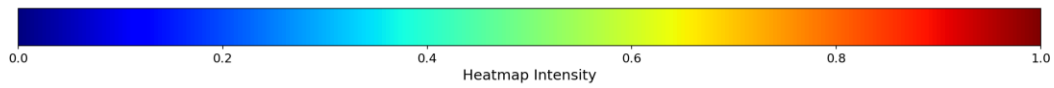
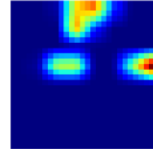
KaNet5: L2 Norm
Predicted Class: 6



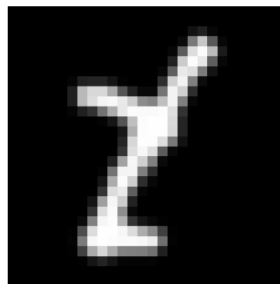
KaNet5: No Norm Heatmap



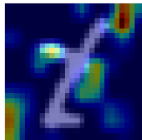
KaNet5: L2 Norm Heatmap



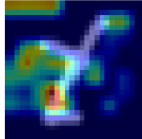
Original Image - Correct Class: t



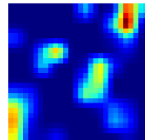
LeNet5: No Norm
Predicted Class: y



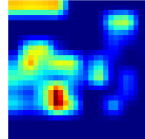
LeNet5: L2 Norm
Predicted Class: t



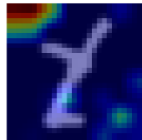
LeNet5: No Norm Heatmap



LeNet5: L2 Norm Heatmap



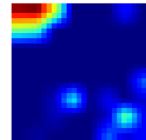
KaNet5: No Norm
Predicted Class: t



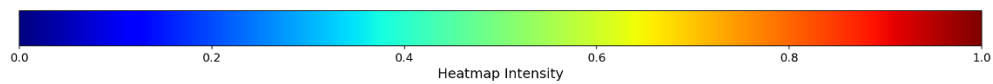
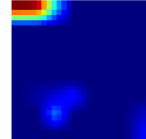
KaNet5: L2 Norm
Predicted Class: t



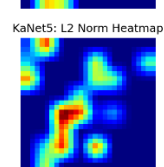
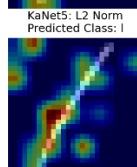
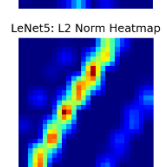
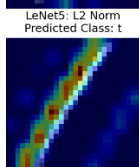
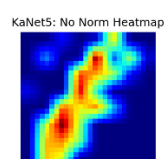
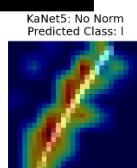
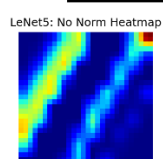
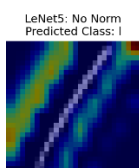
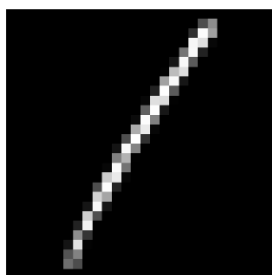
KaNet5: No Norm Heatmap



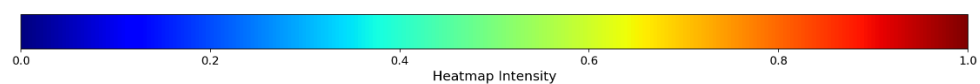
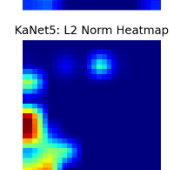
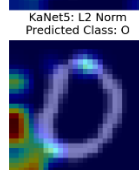
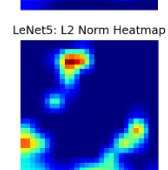
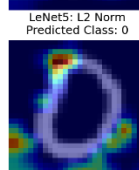
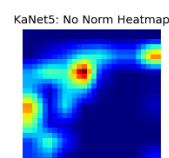
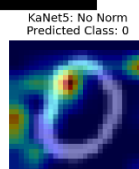
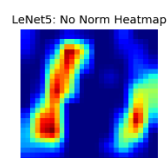
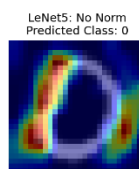
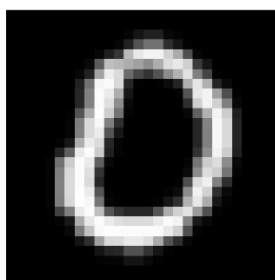
KaNet5: L2 Norm Heatmap



Original Image - Correct Class: l



Original Image - Correct Class: 0



Statistiche sulle Heatmap

I CSV contengono valori medi, varianze e massimi dell'intensità Grad-CAM. Nei casi classificati correttamente l'intensità risultano più concentrate; nei casi errati l'intensità diminuisce. ConvKAN mostra maggiore varianza di attivazione.

=== LeNet5: No Norm ===

	Description	mean_intensity	variance_intensity	max_intensity
0	Train correct	5.440341	167.731094	91.789970
1	Train incorrect	5.334188	156.757812	85.210648
2	Test correct	5.440900	167.764359	91.807999
3	Test incorrect	5.347645	158.631439	86.198540

=== KaNet5: No Norm ===

	Description	mean_intensity	variance_intensity	max_intensity
0	Train correct	0.536035	427.624939	70.927231
1	Train incorrect	0.561196	367.806702	62.290306
2	Test correct	0.540707	427.426666	70.952354
3	Test incorrect	0.557997	374.891510	63.157063

=== LeNet5: L2 Norm ===

	Description	mean_intensity	variance_intensity	max_intensity
0	Train correct	6.342367	219.083313	103.019150
1	Train incorrect	6.156100	201.739182	95.141525
2	Test correct	6.343664	219.131668	103.083282
3	Test incorrect	6.171502	204.006851	96.152519

=== KaNet5: L2 Norm ===

	Description	mean_intensity	variance_intensity	max_intensity
0	Train correct	0.910028	576.164246	81.579407
1	Train incorrect	0.816573	493.616119	71.456688
2	Test correct	0.915170	575.786743	81.592850
3	Test incorrect	0.814378	503.853394	72.541130

6. Discussione

Confronto tra LeNet5 e ConvKAN

- **LeNet5:**
 - Semplice, interpretazione più lineare.
 - Heatmap stabili e ben focalizzate.
 - Performance solide ma standard.
- **ConvKAN:**
 - Filtri spline complessi, attivazioni più diffuse.
 - Leggero miglioramento prestazionale, costo computazionale maggiore.
 - I filtri non lineari forniscono flessibilità, ma richiedono analisi più approfondite.

La natura spline dei filtri ConvKAN è un punto di forza, ma richiede maggiore cura e interpretazione.

Limiti e Sfide

- E-MNIST è semplice, su dataset più complessi differenze maggiori.
- Necessario tuning di `grid_size` e `spline_order`.
- ConvKAN richiede maggiore sforzo di interpretazione.

Possibili Estensioni

- **Test su CIFAR-10/100**
- **Ulteriori tecniche di interpretazione** (LRP, Integrated Gradients)
- **Pruning e riduzione parametri**
- **Studio sistematico di `spline_order` e `grid_size`**

7. Conclusioni

Sintesi dei Risultati

LeNet5 e ConvKAN hanno prestazioni simili su E-MNIST, ma Grad-CAM rivela differenze nelle strategie di attenzione interna. LeNet5 è più lineare e focalizzata, ConvKAN mostra maggiore varietà di attivazione, coerente con la sua natura spline adattiva. Questo suggerisce che i filtri non lineari interni a ConvKAN consentono alla rete di vedere pattern più complessi.

Bibliografia

- LeCun, Y. et al., "Gradient-Based Learning Applied to Document Recognition," Proceedings of the IEEE, 1998.
- Cohen, G. et al., "EMNIST: Extending MNIST to handwritten letters," arXiv:1702.05373, 2017.
- Selvaraju, R. R. et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," ICCV 2017.
- Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). Kolmogorov-Arnold Networks (KANs): Promising alternatives to Multi-Layer Perceptrons (MLPs). Preprint. Retrieved from <https://arxiv.org/abs/2404.19756>
- Bodner, A. D., Tepsich, A. S., Spolski, J. N., & Pourteau, S. (2024). Convolutional Kolmogorov-Arnold Networks. *arXiv preprint arXiv:2406.13155*. <https://doi.org/10.48550/arXiv.2406.13155>