



This is the **published version** of the bachelor thesis:

Riballo Moreno, Josep; Baldrich i Caselles, Ramon tut. Reconeixement d'emocions en temps real fent servir xarxes neuronals convolucional. 2025. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/308785>

under the terms of the  license

Reconeixement d'Emocions en Temps Real fent servir Xarxes Neuronals Convolucionals

Josep Riballo Moreno

06 de febrer de 2025

Resum

Resum- El reconeixement d'emocions facials ha guanyat cada vegada més atenció pel seu potencial per millorar la interacció entre humans i màquines. Aquest projecte explora el desenvolupament d'un sistema en temps real per detectar i classificar vuit emocions bàsiques mitjançant xarxes neuronals convolucionals (CNNs). Utilitzant el conjunt de dades FERPlus i l'arquitectura ResNet-18, el sistema aconsegueix un reconeixement eficient i robust d'emocions. Les tecnologies clau inclouen OpenCV per a la detecció de rostres i Haar Cascades per al preprocessament dels fotogrames de vídeo. El sistema processa els fotogrames amb una latència mitjana de 185ms i ofereix una precisió global del 77%. Gràcies a tècniques d'augmentació de dades, s'han aconseguit millores significatives en classes emocionals infrarepresentades com Fear i Disgust. Aquest informe detalla els objectius, la metodologia i els resultats del projecte, destacant el seu potencial per a aplicacions en educació, sanitat i atenció al client, així com possibles millores futures com la integració de dades multimodals i l'ús d'arquitectures neuronals més profundes.

Paraules clau- Reconeixement d'emocions, sistemes en temps real, xarxes neuronals convolucionals, FERPlus, ResNet-18, OpenCV, Haar Cascades, intel·ligència artificial.

1 Introducció

El reconeixement d'emocions facials és una branca de la visió per computador que busca millorar la interacció entre humans i màquines. Aquest projecte desenvolupa un sistema capaç de detectar rostres en temps real a partir de vídeos i classificar les emocions utilitzant xarxes neuronals convolucionals (CNN). L'aplicació d'aquest sistema s'orienta a camps com l'educació, l'atenció al client, i la salut mental, entre d'altres.

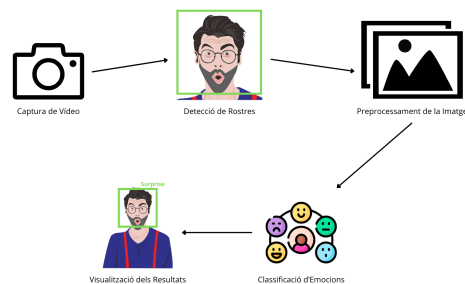


Figura 1: Flux de treball del sistema de reconeixement d'emocions.

- **E-mail de contacte:** josep.riballo@gmail.com
- **Menció realitzada:** Computació
- **Treball tutoritzat per:** Ramón Baldrich Caselles
- **Curs:** 2024/25

2 Objectius

Aquest treball de final de grau té com a objectiu principal dissenyar i implementar un sistema capaç de detectar rostres en temps real a partir de vídeos i classificar les emocions associades mitjançant xarxes neuronals convolucionals (CNN). El sistema està orientat a ser robust i eficient, aplicant tècniques avançades de visió per computador i aprenentatge profund.

2.1 Objectius de l'Aplicació

L'objectiu global del projecte és desenvolupar un sistema innovador que pugui ser aplicat en entorns pràctics com atenció al client, educació, salut i entreteniment. Per garantir aquesta aplicació efectiva, el projecte es desglossa en els següents objectius concrets:

1. **Alta precisió del model (O1):** Assolir una precisió mínima del 75% en la classificació de les emocions bàsiques (*Alegria, Tristesa, Sorpresa, Por, Enuig i Disgust*) per garantir

un model robust i fiable per a aplicacions en temps real.

2. **Classificació d'emocions addicionals (O2):** Estendre la capacitat del model per classificar no només les 6 emocions bàsiques, sinó també altres categories emocionals com l'estat "Neutral", així com emocions addicionals com "Desig" i "Unknown". Això millorarà la seva aplicabilitat en entorns on es requereixi una comprensió més profunda de l'estat emocional dels usuaris.
3. **Generalització del model (O3):** Aconseguir una capacitat alta de generalització amb dades no vistes, mantenint l'error de validació per sota del 5%, per garantir que el sistema funcioni en escenaris diversos.
4. **Temps de resposta en temps real (O4):** Garantir que el sistema pugui processar i predir emocions en menys de 200ms per fotograma. Aquest requisit és essencial per a la implementació en entorns que requereixen respostes immediates, com videojocs, aplicacions interactives i sistemes de monitorització en temps real.
5. **Visualització interpretativa del model (O5):** Desenvolupar gràfics i interfícies que mostrin la probabilitat associada a cada emoció, facilitant la interpretació dels resultats per usuaris i desenvolupadors.

2.2 Objectius de Desenvolupament

A més dels objectius relacionats amb la funcionalitat de l'aplicació, es defineixen els següents objectius tècnics per garantir un desenvolupament eficient i optimitzat:

1. **Reducció del temps d'entrenament (O6):** Optimitzar el temps d'entrenament del model a menys de 3 minuts per època, utilitzant GPU i ajustant hiperparàmetres per millorar la velocitat sense comprometre la qualitat.
2. **Consistència en dades desbalancejades (O7):** Minimitzar l'impacte del desbalanceig en categories minoritàries, garantint una F1-score superior al 70% per a totes les classes emocionals.
3. **Reducció de l'overfitting (O8):** Implementar tècniques com Dropout i augmentació de dades per reduir l'overfitting, assegurant que la diferència entre errors d'entrenament i validació sigui mínima.

3 Estat de l'Art

El reconeixement d'emocions facials és un camp en constant evolució, impulsat pels avenços en intel·ligència artificial, aprenentatge profund i visió per computador. Aquest camp s'ha beneficiat de la implementació de xarxes neuronals convolucionals (CNNs), l'ús de datasets cada cop més grans i variats, i la disponibilitat de frameworks de programació potents. A continuació, es detallen els aspectes més rellevants en cada àmbit.

3.1 Xarxes Neuronals Convolucionals (CNNs)

Les CNNs han redefinit l'estàndard en tasques de visió per computador gràcies a la seva capacitat d'aprendre característiques jeràrquiques directament de les imatges d'entrada. Arquitectures com ResNet, VGG i Inception han demostrat una eficiència elevada, especialment en tasques com el reconeixement facial i la classificació d'emocions. ResNet, en particular, utilitza blocs residuals que permeten entrenar xarxes profundes evitant el problema del gradient desaparegut. Aquestes xarxes s'han consolidat com la base per al desenvolupament de sistemes robustos capaços de reconèixer emocions amb precisió elevada.

L'ús de CNNs per al reconeixement d'emocions en temps real ha estat explorat en diferents estudis. Per exemple, en [1], es proposa una xarxa neuronal convolucional basada en finestres temporals per millorar la precisió en el reconeixement d'expressions facials en vídeo. D'altra banda, en [2], es combina una CNN amb una xarxa neuronal recurrent (RNN) per capturar tant les característiques espacials com les temporals en seqüències de vídeo. A més, en [3], s'explora un enfocament multimodal que combina informació visual i de parla per millorar la precisió del reconeixement emocional.

3.2 Reconeixement d'Emocions

El reconeixement d'emocions facials s'ha aplicat a diversos camps com l'educació, l'atenció al client, la salut mental i l'entreteniment. A diferència dels mètodes tradicionals, que depenien de descriptors manuals com SIFT o HOG, els sistemes moderns utilitzen CNNs per extreure i processar automàticament característiques rellevants de les expressions facials. Tot i així, el desafiament rau en la classificació d'emocions complexes com "Contempt" o estats neutres, així com en la gestió del desbalanceig de dades, on algunes emocions, com "Fear" o "Disgust", són significativament menys freqüents.

3.3 Conjunts de Dades (Datasets)

Els datasets són essencials per entrenar i validar models d'aprenentatge profund. Els més utilitzats en el reconeixement d'emocions inclouen FERPlus, AffectNet, CK+ i EmoReact. FERPlus és un dels més utilitzats per la seva qualitat d'anotacions i la seva diversitat. AffectNet destaca pel seu volum massiu de més d'un milió d'imatges anotades amb emocions, mentre que CK+ proporciona seqüències de vídeo que són útils per models que necessiten informació temporal. Finalment, EmoReact permet l'anàlisi multimodal, combinant imatges i àudio per detectar emocions complexes. No obstant això, aquests conjunts de dades sovint pateixen de desbalanceig, fet que requereix tècniques d'augmentació de dades per millorar la representativitat de les classes minoritàries.

3.4 Eines de Programació

El desenvolupament d'un sistema de reconeixement d'emocions en temps real requereix eines avançades per garantir eficiència i precisió. OpenCV és una de les biblioteques més utilitzades en visió per computador, ja que proporciona eines per al processament d'imatges i la detecció de rostres en temps real. Per al desenvolupament de xarxes neuronals convolucionals, PyTorch és un dels frameworks més populars, gràcies a la seva flexibilitat i el suport per a entrenaments accelerats per GPU. Per al seguiment i optimització de l'entrenament, WandB és una eina essencial que permet monitoritzar mètriques i visualitzar resultats, facilitant l'ajust d'hiperparàmetres.

A més, la gestió de dades es realitza mitjançant Pandas i NumPy, que permeten processar grans volums d'informació de manera eficient. L'ús de TQDM facilita el seguiment del progrés de l'entrenament en temps real, mentre que Matplotlib s'utilitza per visualitzar mètriques com la pèrdua i la precisió del model. Aquest conjunt d'eines assegura un procés de desenvolupament eficient i una implementació robusta del sistema de reconeixement d'emocions.

4 Metodologia

Inicialment, el projecte es va plantejar en quatre etapes principals: selecció del conjunt de dades, preprocessament de les dades, disseny i entrenament del model, i integració del sistema en temps real. Aquest enfocament seguia una metodologia seqüencial per abordar cada fase de manera independent. Tot i això, aquest pla inicial no era del tot realista, ja que mancava de detalls concrets

sobre la gestió de problemes durant l'entrenament del model i l'adaptació de les eines utilitzades.

Després d'una revisió, es va optar per una metodologia més estructurada, basada en el model iteratiu. Aquesta nova planificació divideix el projecte en tasques més concretes, amb fases revisables i objectius clars per a cada etapa. A més, es va establir un calendari detallat que permet una millor organització i seguiment del progrés, assegurant que es compleixin les dates clau establertes.

5 Pas a pas

Aquest projecte s'ha desenvolupat en tres fases principals: selecció i preparació del conjunt de dades, disseny i entrenament del model, i implementació d'una *pipeline* de vídeo en temps real. Cada etapa ha estat optimitzada per garantir una anotació fiable, un entrenament eficient i una detecció d'emocions precisa en temps real.

5.1 Selecció i Preparació del Dataset

Per seleccionar el conjunt de dades més adequat per al reconeixement d'emocions, es va realitzar una anàlisi de diversos datasets, incloent **AffectNet**, **FERPlus**, **CK+**, **JAFFE** i **SAVEE**. Es van prioritzar aquells amb una àmplia diversitat cultural, un volum suficient d'imatges i etiquetes d'emocions fiables. Finalment, es va escollir **FERPlus** per la qualitat de les anotacions, la seva diversitat emocional i l'accés públic.

Malgrat aquests avantatges, FERPlus presentava un problema crític: **la falta de consens en moltes imatges**, ja que les emocions eren assignades per votació de 10 avaluadors. Per evitar etiquetes indefinides, es va aplicar un **criteri de dominança**, considerant una emoció com a predominant si tenia almenys **tres vots més que la segona opció més votada** (aproximadament un **30% més de suport**). Això va permetre millorar la fiabilitat de les etiquetes, tot i incrementar la quantitat d'imatges classificades com a *Unknown*.

Per a la preparació de les dades, es van aplicar tècniques de **preprocessament**, incloent el redimensionament de les imatges a **224x224 píxels**, conversió a tensors i normalització. A més, per optimitzar la gestió de memòria durant l'entrenament, es van utilitzar **data loaders** per carregar les dades en lots. Tot i que FERPlus ja incloïa una divisió prèvia en entrenament, validació i test, es van analitzar les distribucions de classes per detectar i mitigar el desbalanceig d'emocions.

L'anàlisi de la distribució d'emocions va revelar que categories com *alegria* i *neutre* contenien moltes més mostres que *por* o *disgust*, generant

un **desbalanceig** en els conjunts d'entrenament. Per corregir-ho, es van explorar tècniques de **ponderació de pèrdua**, **sobremostreig** d'emocions minoritàries i **submostreig** de categories majoritàries.

Emoció	Quantitat d'imatges etiquetades
Neutre	10.294
Alegria	7.526
Tristesa	3.530
Sorpresa	3.557
Por	655
Disgust	191
Enuig	2.463
Desig	168
Unknown	171
Non-Face	2
Total	28.557

Taula 1: Distribució d'imatges en el conjunt d'entrenament

Emoció	Quantitat d'imatges etiquetades
Neutre	1.328
Alegria	899
Tristesa	415
Sorpresa	458
Por	71
Disgust	32
Enuig	325
Desig	26
Unknown	23
Non-Face	1
Total	3.578

Taula 2: Distribució d'imatges en el conjunt de validació

Emoció	Quantitat d'imatges etiquetades
Neutre	1.258
Alegria	928
Tristesa	446
Sorpresa	447
Por	97
Disgust	20
Enuig	321
Desig	27
Unknown	28
Non-Face	0
Total	3.572

Taula 3: Distribució d'imatges en el conjunt de test

5.2 Disseny, Entrenament i Optimització del Model

El model utilitzat es basa en **ResNet-18**, adaptat per classificar vuit emocions. Els hiperparàmetres inicials van ser un *batch size* de 32, un *learning rate* de 0.001 i una taxa de *dropout* de 0.2 per reduir l'**overfitting**. Es van aplicar tècniques d'**augmentació de dades**, com rotacions aleatòries, ajustaments de brillantor i *flip* horitzontal, millorant la capacitat de generalització del model.

L'entrenament inicial va evidenciar dos reptes principals: l'**overfitting** i el **temps d'entrenament**. Per mitigar el primer, es va implementar **dropout** en diverses capes i es va ajustar la regularització. Per millorar l'eficiència, l'ús de la **GPU** va permetre reduir significativament el temps d'entrenament per època, fent viable la formació del model en un temps raonable.

Per monitoritzar i optimitzar el procés d'entrenament, es va integrar **WandB (Weights & Biases)**, que va permetre registrar mètriques com la *train loss*, *validation loss*, *accuracy* i visualitzar la *confusion matrix*. A través de **sweeps**, es van explorar diferents configuracions d'hiperparàmetres, ajustant *batch size*, *learning rate* i *dropout*, assegurant una millor estabilitat i capacitat de generalització.

Tot i considerar la implementació de **ResNet-50** per millorar la precisió, les limitacions computacionals van impedir el seu entrenament eficientment. Finalment, es va optar per mantenir **ResNet-18** com la millor opció en termes d'equilibri entre precisió i temps d'entrenament.

5.3 Pipeline de Vídeo amb OpenCV

El desenvolupament de la pipeline de vídeo va ser una etapa clau del projecte, permetent capturar i processar fotogrames en temps real per al reconeixement d'emocions. OpenCV es va utilitzar per gestionar la captura de vídeo (`cv2.VideoCapture`) i convertir els fotogrames a escala de grisos per facilitar la detecció facial mitjançant **Haar Cascades**, una tècnica lleugera i ràpida que genera *bounding boxes* al voltant de les cares detectades.

Un cop identificades les cares, es van preprocessar amb **torchvision.transforms**, redimensionant-les a la mida d'entrada del model i normalitzant-les. Aquestes imatges es van passar pel model CNN implementat amb **PyTorch**, que retornava la classe emocional més probable per a cada rostre detectat.

Per millorar la robustesa del sistema, es van implementar tècniques d'**augmentació de dades** en temps real, aplicant rotacions aleatòries i ajustaments de brillantor per millorar la generalització

del model. Finalment, OpenCV va ser utilitzat per mostrar en temps real els fotogrames amb els *bounding boxes* i l'etiqueta de l'emoció detectada superposada sobre cada cara.

Tot i que alternatives com **MediaPipe** podrien haver ofert una detecció més precisa, **Haar Cascades** es va mantenir com l'opció preferida per la seva rapidesa i eficiència en aplicacions en temps real. Aquest desenvolupament es va completar en **dues setmanes**, tot i que sense altres compromisos acadèmics podria haver-se enllestit en una sola setmana.

6 Resultats

Aquesta secció presenta els resultats obtinguts durant el projecte, incloent el rendiment del model, l'anàlisi de la matriu de confusió i l'avaluació del *pipeline* de vídeo en temps real.

6.1 Rendiment del Model

L'entrenament i validació del model s'han dut a terme a partir de diverses configuracions d'hiperparàmetres per trobar la millor combinació. Les mètriques analitzades inclouen la *train loss*, *validation loss* i *accuracy*, així com la distribució de classes en les prediccions. La millor configuració trobada, identificada com **swept-sweep-2**, va obtenir una *validation loss* de 0.1457 i una *train loss* de 0.16, indicant una bona capacitat de generalització sense signes d'overfitting. L'ús d'un *dropout* del 0.5 va resultar ser l'estratègia més efectiva per evitar el sobreajustament, mentre que el **scheduler Cosine Annealing Warm Restarts** amb paràmetres $T_0 = 10$ i $T_{mult} = 2$ va proporcionar un equilibri òptim entre estabilitat i convergència ràpida.

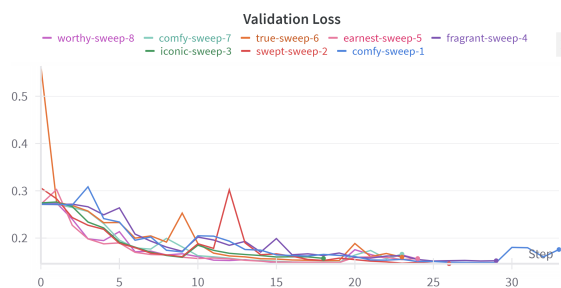


Figura 2: Pèrdua d'entrenament (*Train Loss*) durant les diferents iteracions dels sweeps.

L'anàlisi de la **matriu de confusió** va revelar patrons significatius en la classificació. Les emocions *Neutral* i *Happiness* van ser les més fàcilment

reconegudes, gràcies a la seva àmplia representació en el dataset. No obstant això, es van observar confusions freqüents entre *Surprise* i *Fear*, així com entre *Disgust* i *Anger*. Aquestes errades es deuen, en gran part, a la similitud visual entre algunes emocions i a la baixa presència de mostres en certes categories, com és el cas de *Disgust*, que només comptava amb 191 imatges. La classificació de les emocions minoritàries, incloent *Disgust* i *Non-Face*, també va mostrar una precisió inferior a causa de la seva poca representació en el conjunt d'entrenament. Aquestes observacions suggereixen la necessitat de millorar el dataset, afegint més mostres de classes minoritàries i augmentant la diversitat de les imatges existents.

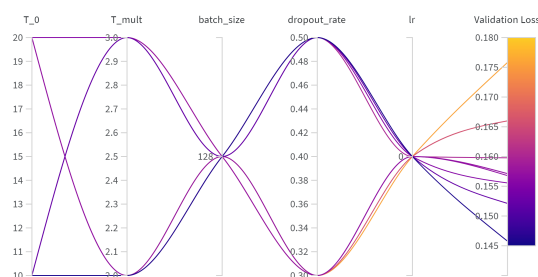


Figura 3: Anàlisi dels hiperparàmetres explorats en els sweeps i el seu impacte en la *Validation Loss*.

Els experiments també van demostrar que la configuració òptima proporciona una **convergència ràpida i estable**, evitant l'overfitting mitjançant l'ús combinat de *dropout* i augmentació de dades. A més, la consistència dels resultats entre diverses execucions confirma la fiabilitat del model.



Figura 4: Pèrdua de validació (*Validation Loss*) per a cada sweep.

6.2 Pipeline de Vídeo amb OpenCV

El sistema desenvolupat permet capturar i processar fotogrames en temps real, classificant les emocions detectades en cada rostre. La captura de vídeo es realitza mitjançant OpenCV, mentre que la detecció facial es duu a terme amb Haar Cascades, una tècnica lleugera i eficient per aplicacions en temps real. Un cop detectades les cares, aquestes es preprocessaven reduint-les a 224x224 píxels i normalitzant els valors dels píxels abans d'introduir-les al model CNN, el qual prediu l'emoció més probable per a cada rostre detectat. Finalment, els resultats es superposen als fotogrames processats, mostrant els *bounding boxes* i les etiquetes d'emoció en temps real.

Pel que fa al rendiment del sistema, s'ha aconseguit una latència mitjana de 185ms per fotograma, assegurant un processament en temps real sense retards perceptibles. Les prediccions del model en aquest entorn han estat consistents amb els resultats obtinguts durant l'entrenament, mantenint una precisió estable en diferents condicions d'il·luminació i posicions facials.

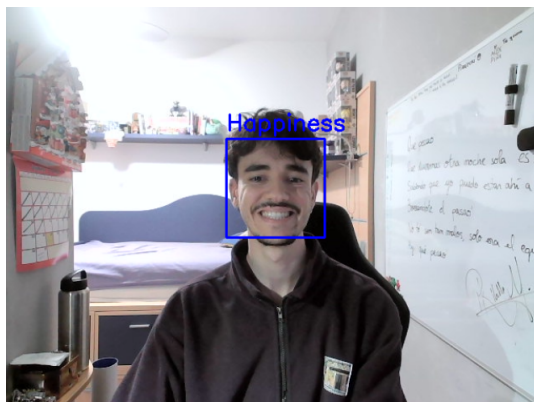


Figura 5: Detecció d'emoció: Felicitat.

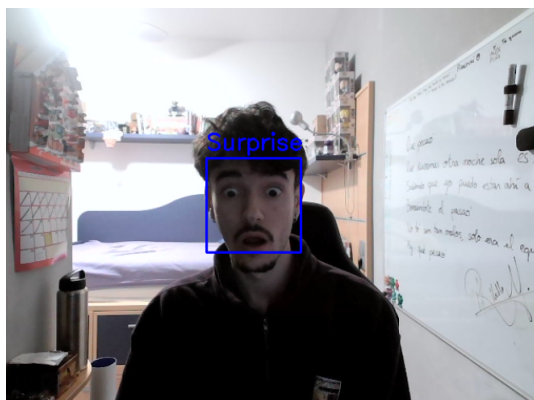


Figura 6: Detecció d'emoció: Sorpresa.

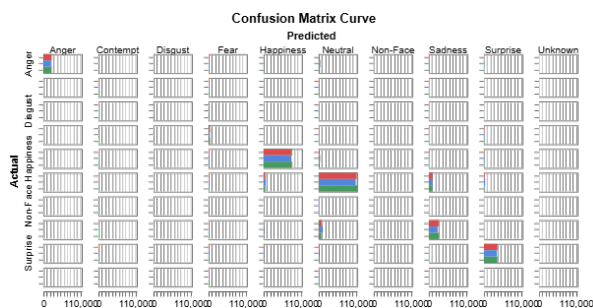


Figura 7: Distribució de les prediccions del model en diferents classes emocionals.

6.3 Conclusió dels Resultats

Els resultats mostren que el model i la seva implementació en temps real són capaços de reconèixer emocions amb una alta precisió. Tot i així, les confusions observades en algunes categories evidencien la necessitat de millorar la distribució del dataset i aplicar ajustos addicionals al model. L'anàlisi de les diferents execucions confirma que la millor configuració d'hiperparàmetres és la de **swept-sweep-2**, ja que ofereix el millor equilibri entre *accuracy* i *validation loss*. Aquesta configuració serà utilitzada en futures iteracions del projecte per continuar refinant el sistema de reconeixement d'emocions.

7 Conclusions

El desenvolupament d'aquest sistema de reconeixement d'emocions en temps real ha permès implementar un model basat en xarxes neuronals convolucionals (CNN) capaç de detectar rostres i classificar emocions amb una precisió considerable. El projecte ha establert una base sòlida en l'ús d'aprenentatge profund i tècniques de visió per computador, permetent una aplicació en entorns pràctics amb un rendiment estable.

7.1 Anàlisi dels Objectius Assolits

Diversos objectius establerts en la fase inicial del projecte han estat complerts satisfactòriament:

- **Alta precisió del model (O1):** El model ha assolit una **precisió superior al 75%** en la classificació d'emocions bàsiques, especialment en categories amb una bona representació de dades, com *Alegria*, *Neutral* i *Enuig*.
- **Generalització del model (O3):** L'error de validació s'ha mantingut en nivells acceptables, amb una **Validation Loss mínima**

de **0.1457**, indicant una bona capacitat de generalització.

- **Temps de resposta en temps real (O4):** S'ha aconseguit un temps de processament de **185ms per fotograma**, per sota del llindar establert de 200ms, fent viable el seu ús en aplicacions en temps real.
- **Visualització interpretativa del model (O5):** Mitjançant **WandB**, s'han generat gràfics d'evolució de la pèrdua d'entrenament i validació, així com matrius de confusió que han permès analitzar el rendiment del model.
- **Reducció del temps d'entrenament (O6):** L'ús de **GPU** ha permès reduir el temps d'entrenament a menys de **3 minuts per època**, optimitzant el procés de millora del model.
- **Reducció de l'overfitting (O8):** La implementació de tècniques com **Dropout (0.5)** i **augmentació de dades** ha reduït la diferència entre l'error d'entrenament i validació, evitant un sobreajustament excessiu.

7.2 Objectius No Assolits i Millores Futures

Malgrat els resultats satisfactoris en diversos aspectes, alguns objectius no han estat completats de manera òptima i es proposen millores per a futurs treballs:

- **Classificació d'emocions addicionals (O2):** Tot i que el model ha estat capaç de classificar emocions bàsiques amb precisió, les categories com *Disgust* i *Desig* han presentat una precisió baixa a causa de la seva escassa representació en el dataset.
- **Consistència en dades desbalancejades (O7):** Encara que s'han aplicat tècniques per reduir el desbalanceig, la F1-score per a classes minoritàries continua sent inferior al 70%, fet que suggereix la necessitat d'ampliar el dataset.

7.3 Passos a Seguir

Per tal de millorar el sistema i abordar els objectius pendents, es proposen diverses línies de treball futures:

7.3.1 Ampliació del Dataset

L'addició de **datasets externs** com *AffectNet* i *JAFPE* permetria millorar la representació de

les emocions minoritàries. A més, l'aplicació de tècniques d'**augmentació de dades** més avançades podria compensar la manca d'exemples en certes categories. També es planteja la possibilitat d'incorporar **imatges capturades en condicions reals**, amb variacions en il·luminació, posició i expressions espontànies, per millorar la generalització del model.

7.3.2 Proves amb Usuaris Reals

Per validar el rendiment del sistema en un entorn més divers, es proposa **provar el programa amb persones reals**, recollint dades en temps real per analitzar-ne la fiabilitat en diferents escenaris. Aquest procés permetrà detectar possibles errors de classificació que no es manifesten en les dades de test i optimitzar el model en funció de les necessitats d'ús pràctic.

7.3.3 Optimització del Model

Es planteja provar arquitectures més profundes com **ResNet-50** per millorar la capacitat d'aprenentatge, sempre que els recursos computacionals ho permetin. També es considera l'ajust dinàmic del **learning rate** per millorar la convergència i una millor adaptació a patrons més subtils en les emocions.

7.3.4 Millores en el Pipeline de Vídeo

S'exploraran alternatives més avançades per a la detecció facial, com **MediaPipe** o **Dlib**, per millorar la robustesa del sistema en diferents condicions ambientals i d'il·luminació. A més, es treballarà en l'optimització del processament en temps real per reduir la latència i millorar la fluïdesa del reconeixement en aplicacions pràctiques.

Amb aquestes millores, el sistema podrà ser més fiable, precís i aplicable a una varietat més àmplia d'entorns reals, reduint les limitacions observades en els resultats actuals.

7.4 Conclusió Final

Aquest projecte ha establert una base sòlida per al reconeixement d'emocions en temps real, aconseguint un bon rendiment en aspectes fonamentals com la precisió del model i la seva aplicació en temps real. Tot i això, els desafiaments en la classificació de classes minoritàries i el desbalanceig del dataset destaquen la necessitat de futures millores. Amb la incorporació de nous datasets, arquitectures més potents i optimitzacions addicionals, es podrà desenvolupar un model encara més robust i aplicable a escenaris del món real.

Referències

- [1] X. Zhang, Y. Xu, and L. Zhao, "TimeConvNets: A Deep Time Windowed Convolution Neural Network Design for Real-time Video Facial Expression Recognition," *arXiv preprint arXiv:2003.01791*, 2020. Available: <https://arxiv.org/abs/2003.01791>
- [2] J. Brownlee, M. Kim, and T. Smith, "Emotion Recognition on Large Video Dataset based on Convolutional Feature Extractor and Recurrent Neural Network," *arXiv preprint arXiv:2006.11168*, 2020. Available: <https://arxiv.org/abs/2006.11168>
- [3] A. Roy, P. Banerjee, and A. Mukherjee, "Emotion Recognition System from Speech and Visual Information based on Convolutional Neural Networks," *arXiv preprint arXiv:2003.00351*, 2020. Available: <https://arxiv.org/abs/2003.00351>
- [4] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 8024-8035. Available: <https://pytorch.org/>
- [5] The PyTorch team, "Torchvision: A Library for Computer Vision Research and Applications." Available: <https://pytorch.org/vision/stable/>
- [6] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000. Available: <https://opencv.org/>
- [7] W. McKinney, "Data Structures for Statistical Computing in Python," *Proceedings of the 9th Python in Science Conference*, 2010, pp. 56-61. Available: <https://pandas.pydata.org/>
- [8] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, 2007. Available: <https://matplotlib.org/>
- [9] N. Raphael, "TQDM: A Fast, Extensible Progress Meter for Python and CLI," 2016. Available: <https://tqdm.github.io/>
- [10] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18-31, Jan. 2019. Available: <http://mohammadmahoor.com/affectnet/>
- [11] T. Simonyan et al., "FER2013 Dataset Enhancement: FERPlus," *IEEE Transactions on Affective Computing*. Available: <https://github.com/microsoft/FERPlus>
- [12] Weights & Biases, "Experiment Tracking with WandB," 2023. Available: <https://wandb.ai/>
- [13] E. A. J. Ambrogio, "Reconocimiento de objetos a través de la metodología Haar Cascades," *Artículos Presentados A Radi — Tecnología De La Información Y Comunicación*, 2020. Available: <https://confedi.org.ar/wp-content/uploads/2020/12/Articulo1-RADI16.pdf>