

Informatika Ingeniaritzako Gradua

Konputazioa

Gradu Amaierako Lana

Bideoetako ekintzen identifikaziorako eta sailkapenerako urratsak

Egilea

Manex Atxa Landa

2023

Informatika Ingeniaritzako Gradua

Konputazioa

Gradu Amaierako Lana

Bideoetako ekintzen identifikaziorako eta sailkapenerako urratsak

Egilea

Manex Atxa Landa

Zuzendaria(k)

Itziar Irigoien Garbizu

Itsaso Rodriguez Moreno

Laburpena

Lan honen helburu nagusia ikusmen artifizialaren bidez eta, zehazki, pertsonen eskeletoen eta horietatik erauzitako atributuen bidez, aparkaleku batean dauden indibiduen ekintzak identifikatzea da.

Horretarako, bideoetako pertsonen eskeletoen erauzketa eta eskeleto horien atributuen erauzketa egin dira, hipotesi batzuk oinarri hartuta. Ondoren, lortutako atributuekin neurona-sare bat entrenatu da. Neurona-sarean exekuzio desberdinak eginda, ereduaren portaera aztertu da aldagai batzuen arabera, hala nola, kameratik urruti dauden pertsonak kontuan hartzea ala ez, lortutako eskeletoen datuak leuntzea ala ez, edota eskeletoekin soilik lortzen diren emaitzak atributuekin hobetzen diren ala ez.

Aparkaleku batean, hau da, testuinguru "arriskutsu" batean, identifikatu diren eta auto autonomoaren ikuspegipean arriskua areagotu dezaketen ekintzak lau dira: korrika ibiltzea, mugikorrean txateatzea, telefonoz hitz egitea eta atzera begira egotea. Bosgarren ekintza bat ere identifikatu da, baina, arriskua sortzen ez duen arren, gainontzekoak baztertzeko beharrezkoa dena: oinez ibiltzea. Hori burutzeko, 84.650 instantziatako datu-base bat erabili da, eta instantzia bakoitzak *frame* bakoitzeko pertsona bat eta bere ekintza jasotzen ditu.

Lortutako emaitzek erakusten dute trebatutako eredia zenbateraino den gai ekintzak identifikatzeko, eta ekintzen identifikazio-faseetan zein datu-tratamendu diren erabilgarriak. Beraz, batetik, eredia mundu errealerara aplikatu daitekeen ala ez ondorioztatu da, eta, bestetik, ekintzen sailkapeneko prozesuetan aintzat hartu beharreko alderdi batzuk ondorioztatu dira.

Gaien aurkibidea

| | |
|----------------------------------------------------------------------------------------------------------|------------|
| Laburpena | i |
| Gaien aurkibidea | iii |
| Irudien aurkibidea | v |
| Taulen aurkibidea | vii |
| 1 Sarrera | 1 |
| 2 Proiektuaren Helburuen Dokumentua | 3 |
| 3 Oinarri teorikoak | 7 |
| 3.1 Artearen egoera | 7 |
| 3.2 Irudietatik ezaugarriak ateratzeko teknikak | 8 |
| 3.2.1 Gradiente Orientatuen Histograma | 8 |
| 3.2.2 Eskala-Inbarianteko Ezaugarri Transfomazioa (<i>Scale-invariant feature transform</i>) | 14 |
| 3.2.3 Neurona-sare konboluzionalak (Convolutional Neural Networks edo CNN) | 17 |
| 3.2.4 Eredu Bitar Lokalak (<i>Local Binary Patterns</i> edo LBP) | 20 |
| 3.2.5 Eskeletoak | 21 |
| 3.2.6 Kaxa inguratzailleak (<i>Bounding Box</i> edo BBox) | 24 |
| 3.3 Sailkapen gainbegiraturako teknikak | 24 |
| 3.3.1 <i>Multilayer Perceptron</i> neurona-sareak (MLP) | 25 |
| 3.3.2 Neurona-sare konboluzionalak (<i>Convolutional Neural Networks</i> edo CNN) | 26 |

| | | |
|----------|--------------------------------------------------------|-----------|
| 4 | Egindako esperimentazioa eta lortutako emaitzak | 29 |
| 4.1 | Erabilitako datu-multzoa | 30 |
| 4.2 | Aurreprozesaketa | 31 |
| 4.2.1 | Eskeletoak eraztea | 31 |
| 4.2.2 | Eskeletoen leunketa | 33 |
| 4.2.3 | Atributuak lortzea | 35 |
| 4.3 | Sailkapena | 38 |
| 4.4 | Esperimentazioa | 40 |
| 4.5 | Emaitzak | 41 |
| 5 | Ondorioak | 45 |
| | Bibliografia | 47 |

Irudien aurkibidea

| | | |
|------|-------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Lanaren denbora-banaketa atalen arabera. | 4 |
| 2.2 | Lanaren garapeneko Gantt diagrama. | 5 |
| 3.1 | Oinarri teorikoetan azalduko diren teknikak eta horien sailkapena. | 8 |
| 3.2 | Irudi baten interpretazioa RGB matrizeetan. | 9 |
| 3.3 | RGB matrizeetatik matrize bakarrera soiltzea. | 9 |
| 3.4 | Matrize soilduetako gradienteak lortzeko pausoak. | 10 |
| 3.5 | Bektoreen sailkapena tarteetan eta histograma lortzea. | 11 |
| 3.6 | Gelaxka desberdinetako histogramen bateratzea, bloke baten histograma lortuz. | 12 |
| 3.7 | Hautatutako blokea desplazatuz histogramak normalizatzea. | 12 |
| 3.8 | Gradiente orientatu normalizatuen bateratzea. | 13 |
| 3.9 | Desfokuratze gausiarra eta diferentzia gausiarra tamaina desberdinetako irudietan. [Huang and Wu, 2020] | 15 |
| 3.10 | X-z markatutako pixela berdez markatutako pixelekin alderatzen da leku potentziala den ala ez zehazteko. | 16 |
| 3.11 | Maximoaren %80 baino handiago diren balioak soilik hartzen dira kontuan. | 17 |
| 3.12 | Neurona-sare konboluzionalen egitura. [Saturn Cloud, 2018] | 18 |
| 3.13 | Ezkituko geruzetan filtroak aplikatzea matrize osoa zeharkatuz. [DimitrisMi, 2023] | 19 |
| 3.14 | Ezkituko geruzetan filtroak aplikatzean erabiltzen eta lortzen diren matrizeak. | 19 |
| 3.15 | <i>Max-poolingen</i> funtzionamendua. | 20 |
| 3.16 | LBP adibide baten lehen pausoa, inguruko zenbakiekin alderatzea. | 21 |
| 3.17 | Matrizetik bektorea eta bektoretik zenbaki hamartarra lortzea. | 21 |
| 3.18 | OpenPose arkitekturaren <i>pipelinea</i> . [Cao et al., 2017] | 22 |
| 3.19 | VGG-19 eredu osoa. [Zheng et al., 2018] | 22 |

| | | |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.20 | OpenPosek eskeletoan identifikatzen dituen artikulazioak. | 23 |
| 3.21 | Ibilgailuetan giltza-puntuen detekzio adibidea. [Cao et al., 2017] | 24 |
| 3.22 | <i>Bounding box</i> identifikazio-adibidea. | 24 |
| 3.23 | Aktibazio-funtzio ezagunenen formulak eta grafikatzak. | 26 |
| 3.24 | Adibiderako sortutako neurona-sarearen irudikapena. | 27 |
| 3.25 | Flattening eta Softmax erabilia egindako sailkapen bat. | 28 |
| 4.1 | Proiektuaren eskema orokorra edo pipelinea. | 29 |
| 4.2 | Ekintza bakoitzaren instantzia-kopurua datu-basean. | 31 |
| 4.3 | IoU kalkuluaren funtzionamendua. [Rosebrock, 2016] | 32 |
| 4.4 | Puntu-sorta baten mediana geometrikoa (horiz). | 34 |
| 4.5 | Leunketa-prozesu oso baten adibidea. | 35 |
| 4.6 | Telefonoz hitz egitean belarri-eskuen arteko distantzia laburtzen da. Canva orrialdetik sortua [can, 2023], adimen artifiziala erabilia. | 36 |
| 4.7 | Telefonoz mezuak bidaltzean, besoetako angeluak ixten dira. Canva orrialdetik sortua [can, 2023], adimen artifiziala erabilia. | 36 |
| 4.8 | Korrika goazenean, hanketako angeluak ixten dira. Canva orrialdetik sortua [can, 2023], adimen artifiziala erabilia. | 37 |
| 4.9 | Eraikitako neurona-sarearen irudikapena. | 39 |
| 4.10 | <i>Train</i> eta <i>test</i> partizioen banaketa. | 40 |
| 4.11 | Konfusio-matrizea. | 42 |
| 4.12 | <i>Train</i> eta <i>test</i> partizioen banaketa berria. | 43 |
| 4.13 | Konfusio-matrizea <i>train</i> handituz. | 44 |

Taulen aurkibidea

| | | |
|-----|-------------------------------------|----|
| 2.1 | Atazetan emandako denbora. | 4 |
| 4.1 | Esperimentazioetako datuak. | 41 |

1. KAPITULUA

Sarrera

Informatika arloa hain zabala izanik, bere baitan hartzen dituen azpigaiak ere mamitsuak izaten jarraitzen dute. Horregatik, testuinguru handia izaten dute informatikako gaiak eta, beraz, baita bideoetako ekintzen identifikazioak ere.

Ekintza-errekonozimendua irudi, argazki edota bideoetan gertatzen diren eragileen ekintzak identifikatzean datza, behaketa eta ingurunean oinarrituz. Gizakiek erraz egiten duten identifikazioa izan arren, ordenagailu bidez automatizatzea lortu nahi izan da 80ko hamarkadatik aurrera, eta informatikako ikerketa alor garrantzitsua bihurtu da gainontzeko zientzian duen aplikazioengatik.

Hainbeste bideo mota eta alor izanik, aspaldidanik egon da identifikazio hauetarako interesa. Medikuntzan [Liu et al., 2022], zaintzan [Jin et al., 2017], gizaki-ordenagailu interakzioetan [Rautaray and Agrawal, 2015] edota bideojokoetan [Park et al., 2005] ere erabiltzen dira ekintza-identifikaziorako teknikak. Esaterako, bideo zaintzaren arloan, jende pilaketak dauden lekuak identifikatzeko erabil daiteke, baina baita ekintza indibidualagoak aztertzeko ere. Orokorrean, ordenagailu bidez mundu errealeko irudiak jaso, prozesatu eta horietatik informazioa ateratzea ahalbidetzen duen teknologiarik ikusmen artifizial deritzogu.

Lan honetan ekintza-errekonozimendua grabatutako bideo erreal batzuen gainean burutu da; zehazki, Irlandako Galwayeko Unibertsitateko Connaught Automotive Research Group ikerketa-taldeak [Connaught Automotive Research (CAR) Group, 2005] utzitako bideoen gainean. Aparkaleku batean kokatutako kamera batek jasotako irudietan, pertsonak indibidualki aztertu dira, eta ekintza indibidual zehatz batzuk identifikatzeko eredu bat garatu da. Zehazki, ereduak aparkaleku batean pertsonak egiten dituzten ekintza batzuk identifikatzen ditu: oinez ibiltzea, korrika ibiltzea, telefonoz hitz egitea, mugikorrean txateatzea eta atzera begiratzea. Honekin, oinezkoentzako arriskutsua den ingurune batean arreta galtzea dakarten ekintzak identifikatzea lortu nahi da, hainbat aplikaziotarako baliagarria izan daitekeelako, hala nola, egoera arriskutsuak saihesteko. Oinez ibiltzea, bere horretan arriskutsua ez den arren, gainontzeko ekintzetatik bereizteko identifikatu da.

Helburu nagusi horren barruan, hainbat azpiatal txikiago burutu dira. Pertsonen eskele-

toetatik lortu daitezkeen atributu edo ezaugarri sinpleak eta horien konbinaketak lortzen saiatu da, sailkapenean esanguratsuak izan daitezkeenak. Horrez gain, eskeletotik ateratako atributu hauek sailkapenerako erabiltzea eta eskeletoa bera erabiltzea alderatu ditugu, atributu sinpleekin eskeleto osoa erabiltzean bezain emaitza esanguratsuak lortzeko. Gainera, eskeletoak parametro batzuen arabera leuntzeak zer eragin duen aztertu da, eta baita kameratik urrutiegi dauden pertsonak kontuan ez hartzeak emaitzak hobetzen dituen ere, ondorioztatzeko kamera gehiago jartzeak emaitza hobetuko lukeen. Bestalde, lortutako datuak sailkatzeko modu desberdinak probatu dira, eredu onena lortzeko.

2. KAPITULUA

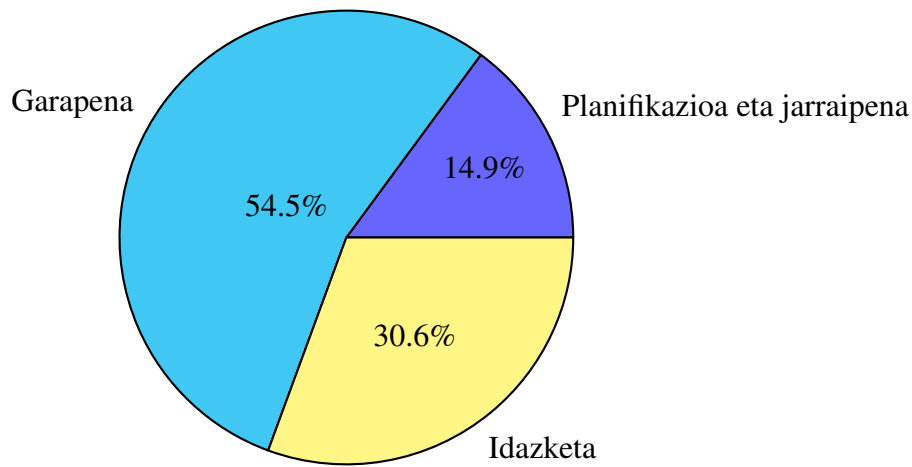
Proiektuaren Helburuen Dokumentua

Lan honen helburu nagusia aparkalekuko bideoetako pertsonen ekintzak identifikatzea da, alegia, ordenagailu bat ekintza horiek identifikatzeko gai izatea lortzea. Helburu nagusi hori burutzeko, lehenengo, bideoetatik informazio esanguratsua lortu da, zehazki, pertsonetatik erauzitako eskeletoetan eta eskeletoetatik atera daitezkeen atribuetan zentratuz. Hori eginda, atributuak beraien artean konparatu ahal izan dira, ekintzak identifikatzeko garaian zein den esanguratsuen jakiteko. Gainera, ondoren, asmatze-tasan eragina izan zezaketen bestelako aldagai batzuk aztertu dira, adibidez, hurbil dauden pertsonak soilik sailkatzean, emaitza hobekia lortzen diren ikustea edota, aurreprozesaketan, lortutako eskeletoen koordinatuak leuntzea -aurreko eta ondorengo posizioak kontuan hartuz-.

Ikusi daitezkeen bezala, asmatze-tasan eragina duten aldagai desberdinak aztertu daitezke, baita eskeletoetatik erauzi daitezkeen atributuak ere. Horregatik, esan daiteke lana progresiboki moldatuz joan daitekeela eta lanean aurrera atributu edo konparazio gehiago gehitu daitezkeela. Hartara, plangintza bera ere progresiboa izan da. Helburu txikiak bakoitza burutu ahala, esperotako ordu kopurura hurbiltzen joan da, eta ia ordu kopurura iristean ez dira atal gehigarri gehiago gehitu, beti ere, lan osatu bat lortu dela kontuan hartuta.

Jarraipena eta antolaketa progresiboak izanda ere, ildo jakin bat jarraitu da. Lehenik eta behin, ekintza-errekonozimenduaren inguruan informazioa jaso da, baita gaiaren egungo egoeraz, teknika desberdinez eta horien sailkapen posibleez ere. Horietatik erabilgarri izan zitezkeenak eta jorratu zitezkeenak hautatu ostean, eskuratutako bideoan zentratu da azterketa. Eskeletoak erauzi dira bideoetatik, eta, ondoren, atributuak lortu. Azkenik, neurona-sare bat entrenatu da eta proba desberdin asko egin, hainbat ondoriotara heltzeko. Dokumentu honen idazketarekin amaitu dira helburu guztiak.

Esan bezala, antolaketa progresibo bat izan da, eta pixkanaka gehitu dira atalak; beraz, hasieran ez zegoen esperotako ordu-antolaketa zehatz bat. Halere, [2.1](#) Irudiak jasotzen du amaieran izandako ordu-banaketa atal orokor bakoitzean.

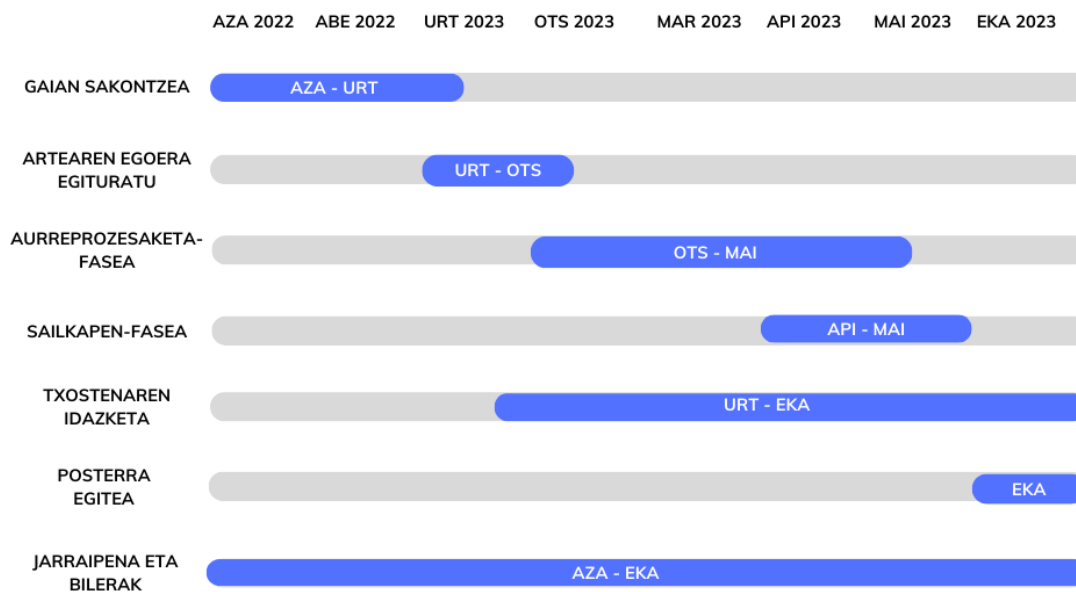


2.1 Irudia: Lanaren denbora-banaketa atalen arabera.

Atal nagusiak atazetan banatu dira eta banaketa hori zehatzago jasotzen da [2.1 Taulan](#) eta [2.2](#) Irudiko Gantt diagraman.

| Atazak | Denbora (ordutan) |
|---------------------------------------------------------|--------------------------|
| Planifikazioa eta jarraipena | 45 |
| Gaian sakontzea eta kokatzea | 35 |
| Jarraipena eta bilerak | 10 |
| Garapena | 165 |
| Artearen egoera egituratzea | 20 |
| Aurreprozesaketa-fasea (leunketa + atributu erauzketa) | 100 |
| Sailkapen-fasea (sarea eraiki + proba desberdinak egin) | 45 |
| Idazketa | 93 |
| Txostenaren idazketa | 90 |
| Posterra egitea | 3 |
| Guztira | 303 |

2.1 Taula: Atazetan emandako denbora.



2.2 Irudia: Lanaren garapeneko Gantt diagrama.

Azkenik, lan hau garatzeko erabili diren tresnak aipatuko dira. Batetik, datu-tratamendu eta exekuzio guztiak Python [Van Rossum and Drake Jr, 1995] lengoiaarekin garatu dira, Spyder [Raybaut, 2009] ingurune integratua erabiliz. Bestetik, honako Pythoneko liburu-tegiak erabili dira: *scikit-learn* [Pedregosa et al., 2011] neurona-sarea eraikitzeke, *pandas* [McKinney et al., 2010] datuak datu-egituretan jaso eta tratatzeko, eta, azkenik, *math* eta *numpy* [Harris et al., 2020] kalkulu eta eragiketak egiteko.

3. KAPITULUA

Oinarri teorikoak

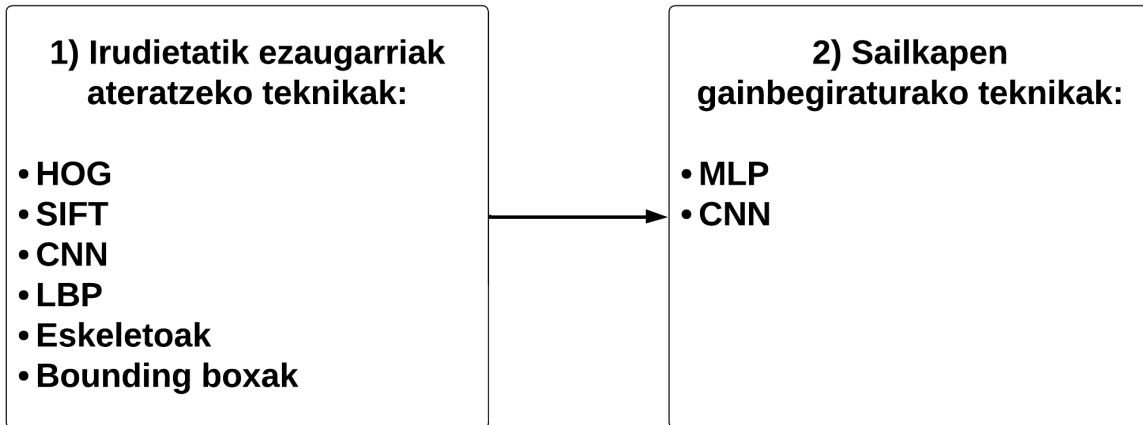
3.1 Artearen egoera

Urteak aurrera joan ahala, teknologia eta ekintza-sailkapenerako teknikak garatuz joan dira. Ikaskuntza sakonaren (*Deep Learning*aren) eztanda baino lehen, eskuz lortutako ezaugarriez baliatutako ereduak erabiltzen ziren. Horien funtsa da irudi geldietatik eta denboran zehar gertatutako aldaketetatik ezaugarriak lortu eta errekonozimendurako erabiltzea. Betiere, errazago da irudi geldietan oinarritzen den informazioa tratatzea; izan ere, denboran zehar gertatzen diren aldaketek informazio gehiago hartzen dute barne, geldiekin alderatuta. Ezaugarriak lortzea esaten denean, hainbat datu-mota desberdin lortzea da helburua, esate baterako, fotograma garrantzitsu, hitz-zaku (*bag-of-words*), interes-puntu eta antzekoak, irudietako informazioa datu-egitura erabilgarrietan jaso behar baita.

Saiakera asko egin dira lortzen diren informazio egiturak tratatzeko eta esanahia emateko. Teknika desberdinak baliatu dira horretarako ere, esaterako, fluxu optikoa (*optical flow*), *Hidden Markov Models* edota, azken aldian, ikaskuntza sakona. Ikaskuntza sakona modu automatikoan datuetatik ikasteko metodoa da, eta, beraz, ikusmen artifizialaren arloan aplikagarria da, azken finean irudiak jaso, prozesatu eta sailkatzean ikasketa prozesu bat baitago. Gradu Amaierako Lan hau esparru horretan kokatzen da, ikusmen artifizialaren arloko prozesuak burutu baitira, bideoetako irudiak aztertu eta informazioa erauziz. Aurrerakuntza nabarmenak egin badira ere, ez dira zehaztasun altuko estandarretara iristen, eta oraindik ataza batzuk ongi funtzionatzeko bidean daude, ikerketa betean.

Jarraian, ekintza-errekonozimendurako teknikak eta horien sailkapen bat ikusiko dira. Batik bat, lan honetan zehar erabili diren teknikak garatuko dira, beste batzuk ere aipatuko diren arren. Arestian esan bezala, bi pauso desberdin nagusitzen dira ekintzen sailkapenean: irudietatik ezaugarriak lortzea eta ezaugarri horietatik ekintzak identifikatzea. Bada, bi ataza horietako bakoitzerako erabilitako teknikak ikusiko dira bakoitza bere azpisailkapenarekin (3.1 Irudia).

OINARRI TEORIKOAK



3.1 Irudia: Oinarri teorikoetan azalduko diren teknikak eta horien sailkapena.

3.2 Irudietatik ezaugarriak ateratzeko teknikak

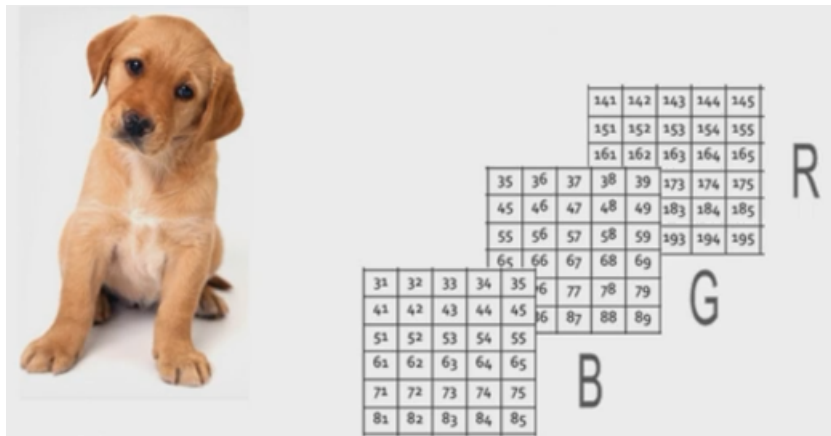
Lehen pausoa *frame* edo irudietatik ezaugarriak lortzean datza. Alegia, jasotako sarrera-irudi bat edota irudi-segida bat izanda, baliagarria izango den bestelako informazioa jasotzeko behar da, kasurako, irudian agertzen diren pertsonen posizioa, abiadura edota esanguratsuak diren bestelako informazio-egiturak (histogramak, bektore karakteristikoak...). Lehen atal honi dagokionez, proiektuan pertsonen eskeletoak eta hauetatik erauzitako atributuak erabili dira, baita *bounding boxak* ere. Eskeletoak eta *bounding boxak*, ordea, ez dira informazioa ateratzeko teknikak bere horretan, hau da, teknika batzuen ondorio dira; adibidez, giltza-puntuak lortzeko tekniken ondorio. Hori dela eta, lehenengo, irudietako zati garrantzitsuak lortzeko teknika batzuk azalduko dira, eta, gero, eskeletoak nola lortzen diren horien konbinazioz.

3.2.1 Gradiente Orientatuen Histograma

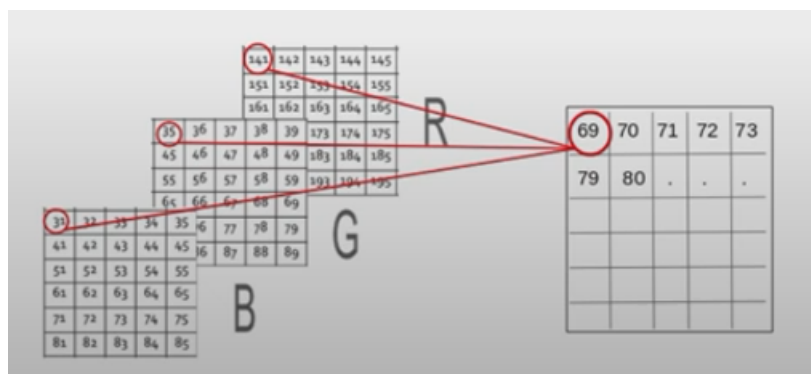
Ezaugarriak lortzeko metodoetako bat Gradiente Orientatuen Histograma (*Histogram of Oriented Gradients* edo HOG) da. Robert K. McCannellek deskribatu zituen lehen kontzeptuak 1986an [McConnell, 1986], eta 1994an erabili zituen Mitsubishi Electric Research Laboratories-ek [Freeman and Roth, 1995]. Hala eta guztiz ere, 2005. urtean hasi zen erabilera zabaltzen Navneet Dalal eta Bill Triggs ikerlariak metodoari buruzko lan

gehigarriaren aurkezpena egin zutenean, zeinetan irudi geldietako oinezkoen identifikazioan zentratu ziren.

Hasteko, ulertu behar da ordenagailu batek nola interpretatzen dituen irudiak. Koloretako irudiak ulertzeko modu ohikoena RGB (gorria, berdea eta urdina) kolore bakoitzeko pixel bakoitzak 0 eta 255 arteko balio bat izatea da (3.2 Irudia). Informazioa murrizte aldera, posizio bereko hiru zenbakien arteko batez bestekoa lortzen da, eta matrize bakarrera murrizten da irudiaren informazioa (3.3 Irudia). Lehenengo, ohikoa da irudia berreskalatzea tamaina aldatuz (32x64 pixeletara, adibidez) eta, ondoren, gelaxkatan (*celletan*) banatzen da (4x4 pixeletako gelaxkatan, kasu).



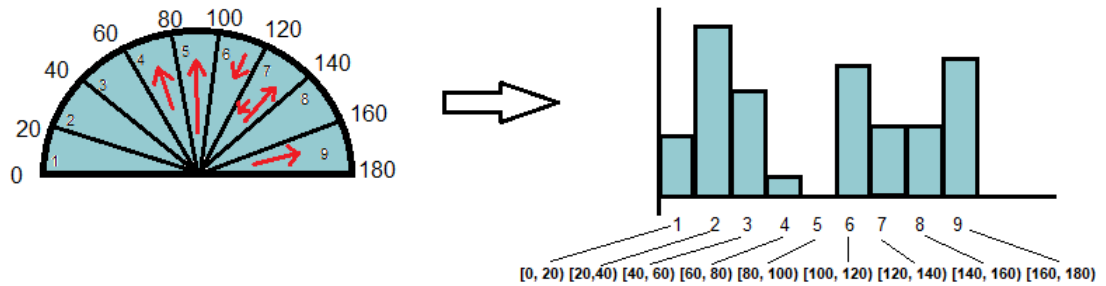
3.2 Irudia: Irudi baten interpretazioa RGB matrizeetan.



3.3 Irudia: RGB matrizeetatik matrize bakarrera soiltea.

Pixel bakoitzeko gradienteak lortu behar dira, eta, horretarako, pixel bakoitzak goian,

nalak egin, eta zatietako bat $[20, 40)$ tartean sartu, eta bestea $[40, 60)$ tartean, izan ere, 36 zenbakia 40tik gertuago dago 20tik baino, eta, horrela, pixka bat errealistago da tarteen banaketa.



3.5 Irudia: Bektoreen sailkapena tartetan eta histograma lortzea.

Esan bezala modu sinpleena balio bakoitza bere osotasunean tarte bakarrean sailkatzea da, eta ez balio bat tarte desberdinetan proportzionalki zatika sartzea. Bada, bektore osoa tarte bakarrean sailkatzeko modu sinple hori erabiliz, ondorengo formulak jasotzen du nola kalkulatu C gelaxka bakoitzeko histograma bakoitza, zehazki, histogramaren tarte bakoitzeko balioak. Azken finean, formulari jasotzen da C gelaxka bateko pixelez pixel aztertuz, momentuan aztertzen ari den histogramako tarte horri baldin badagokio bektorea, g gradientearen gehitu beharko dela.

$$h(k) = \sum_{(x,y) \in C} \omega_k(x,y)g(x,y),$$

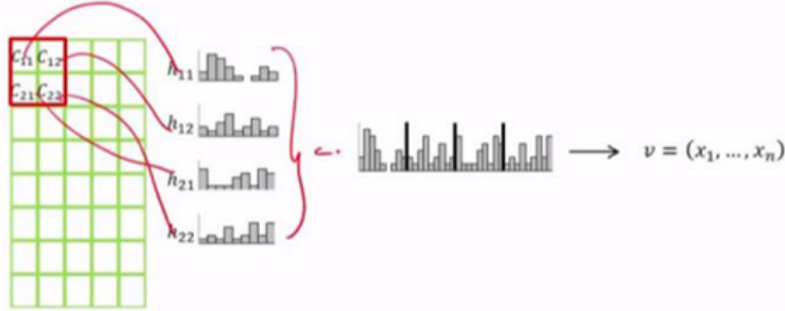
non:

- $h(k)$: histograma k tartean,
- C : histograma kalkulatu nahi den gelaxka,
- $g(x,y)$: gradientearen magnitudea (x,y) posizioan
- $\omega_k(x,y)$, (x,y) pixeleko asignazioaren faktorea, hau da, tarteari dagokion ala ez:

$$\omega_k(x,y) = \begin{cases} 1 & \text{baldin } k \text{ tarteari badagokio} \\ 0 & \text{bestela} \end{cases}$$

Gelaxketako histogramak lortu ostean, bloke izeneko egitura handiagoetan batzen dira.

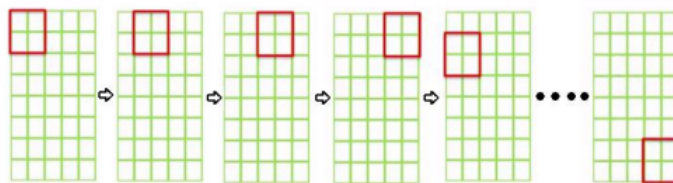
3.6 Irudian ikus daiteke lau gelaxkaz eratutako bloke bat eta nola eratzten den, gelaxka horietako histogrametatik, blokeari dagokion histograma.



3.6 Irudia: Gelaxka desberdinetako histogramen bateratzea, bloke baten histograma lortuz.

Bloke hau desplazatuz joaten da, gelaxka bat eskuinerago mugituz; hortaz, gelaxka bakoitzak bloke batean baino gehiagotan parte hartzen du. Horrela, gelaxketako histogramak izatetik blokeetako histogramak lortzera pasatzen da. Bigarrenik, lortutako histograma berriak normalizatu egiten dira. Hain zuzen ere, irudiaren argiztatzeak gradiente indartsuegiak eman diezazkioke irudiari zati batzuetan; hori dela eta, zati guztietan gradienteek antzeko funtzionamendua izan dezaten eta amaieran lortuko den deskriptoreak errealitatea islatu dezan, histogramak normalizatzea bilatzen da. Hartara, antzeko portaera izango dute gradienteek irudi osoan zehar eta ez dute balio arrarorik hartuko.

Demagun 3.7 Irudiko 4x4 gelaxkako blokeak ditugula, bada, bloke hori pixkanaka desplazatuz eta histogramak normalizatuz joango da, harik eta irudi guztia zeharkatuko den arte.



3.7 Irudia: Hautatutako blokea desplazatuz histogramak normalizatzea.

Esan bezala, v blokeko bektorea emanik, normalizatzeko formula desberdinak erabili daitezke:

L1-norm:

$$\mathbf{f} = \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_1 + e}},$$

L1-sqrt:

$$\mathbf{f} = \sqrt{\frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_1 + e}}},$$

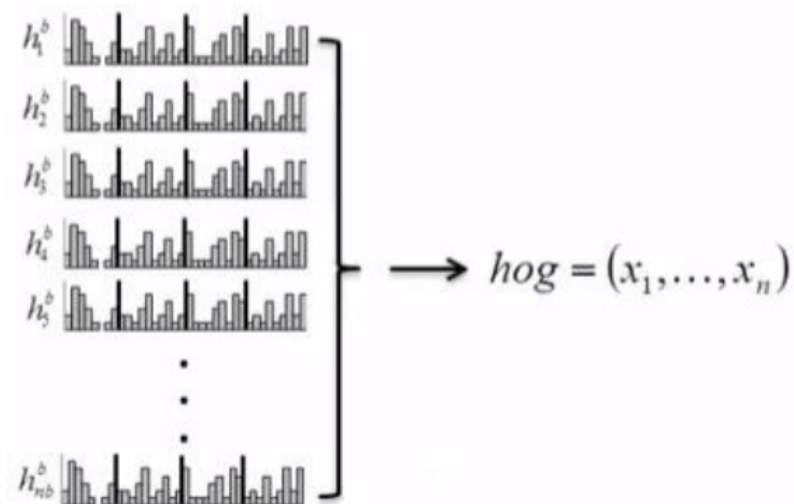
L2-norm:

$$\mathbf{f} = \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2 + e^2}}$$

non:

- \mathbf{v} : bloke bateko histogramak hartzen dituen bektore ez-normalizatua.
- $\|\mathbf{v}\|_k$: bektorearen k-norma, $k=1,2$.
- e : balio txikiko konstantea, zeroa izendatzailean egotea ekiditeko.

Azkenik, bloke guztietan ditugu normalizatutako bektoreak eta deskriptore handi batean bateratzen dira (3.8 Irudia).



3.8 Irudia: Gradiente orientatu normalizatuen bateratzea.

Hau da, hitz gutxitan, teknika honek lortzen duena da, irudi bat emanik, irudi horretako gradiente orientatuak jasotzea deskriptore batean, hainbat sailkapen eta soilketa egin ostean.

3.2.2 Eskala-Inbarianteko Ezaugarri Transfomazioa (*Scale-invariant feature transform*)

Ezaugarriak ateratzeko beste metodo bat ingelesez SIFT siglaz ezagutzen den algoritmoa da, David Lowek asmatua 1999an [Lowe, 1999], eta guztiz definitua eta patentatua 2004an. Ekintzen identifikazioan ez ezik, objektuen identifikazioan, mapaketa robotikoan, 3D modelazioan eta abar erabiltzen da. Modeloaren funtsa HOGen berdina da, hau da, ertzak edo testura-aldaketak bilatzea irudi baten baitan, baina, izenak dioenez, metodo honek irudien eskala-aldaketen eta biraketan aurrean ez du eraldaketarik jasaten.

Bost pauso nagusitan oinarritzen da algoritmoa:

1. Ezaugarriak aurkitzeko leku potentzialak identifikatzea.
2. Giltza-puntuak zehaztea.
3. Giltza-puntuei orientazioak esleitzea.
4. Giltza-puntuak bektore multidimentsionalez deskribatzea.
5. Giltza-puntuen arteko elkartzea.

Ezaugarriak aurkitzeko leku potentzialak identifikatzea

Lehenengo, irudiak desfokuratzen dira desfokuratze gausiarraren bidez. Zehazki, irudia zatikatzen da, eta zati bakoitzeko (x, y) pixel bakoitzean honako kalkulu matematikoa aplikatzen da:

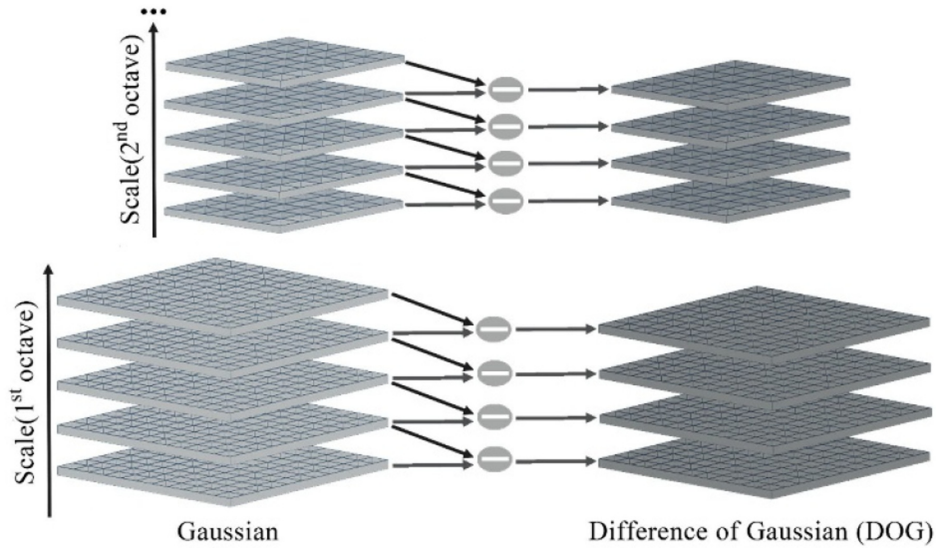
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$I(x, y)$ irudiko pixelak eta $G(x, y, \sigma)$ pixel horri dagokion desfokuratze gausiarraren balioa uztartzen ditu.

G desfokuratze gausiarra aplikatzeko, dagokion balioa ondorengo formularekin kalkulatu da, beti ere, σ eskala parametroa (nolabaiteko desfokuratze kantitatea, zenbat eta handiago sigma orduan eta desfokuratuago) kontuan hartuz:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

Behin tamaina eta desfokuratze desberdineko irudiak lortuta, tamaina bakoitzerako lortutako desfokuratze desberdinekin beste irudi-sorta batzuk sortzen dira. Horretarako, diferentzia gausiarra erabiltzen da, hau da, desfokuratutako irudi desberdinen arteko diferentzia (3.9 Irudia).



3.9 Irudia: Desfokuratze gausiarra eta diferentzia gausiarra tamaina desberdinetako irudietan. [Huang and Wu, 2020]

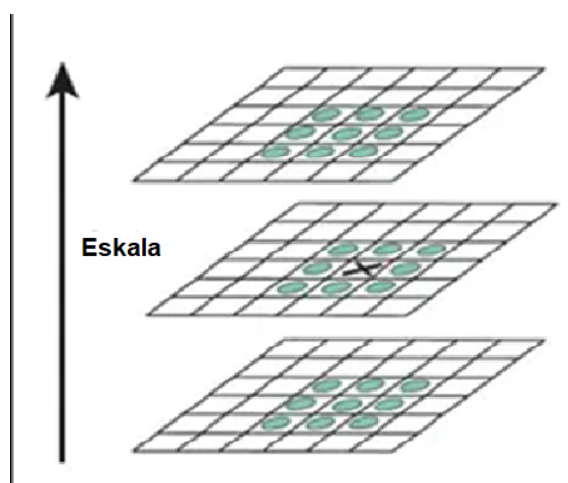
Ondoren, leku-potenzialak izango dira, eskalaz aldatuta ere, lortutako irudi desberdinen artean gehiegi aldatzen ez diren lekuak, gausiarraren laplaziarraren arabera (*Laplacian of Gaussian* edo LoG). LoG operadore honek irudiko intentsitate-aldaketa bortitzak (ertzak) identifikatzen laguntzen du, eta, izenak dioenez, erabili berri den lausotze gausiarrak gain, filtro laplaziarra erabiltzen du; izan ere, laplaziarra zaratarekiko sentikorra denez, erabili den lausotzeak laguntzen dio. $I(x,y)$ intentsitate-balioak dituen irudi bateko pixelaren $L(x,y)$ laplaziarra honela lortzen da:

$$L(x,y) = \frac{\delta^2 I}{\delta x^2} + \frac{\delta^2 I}{\delta y^2}$$

Beraz, orain arteko bi aldaketak nahastuta, azaldu berri den gausiarraren laplaziarra lortzen da:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

Horrela, pixel bakoitza bere inguruko zortziekin eta lausotze desberdina duten goiko eta beheko kapetako bederatziekin alderatzen da, eta, maximo edo minimo lokal bat aurkitu bada, leku potentzialtzat hartuko da (3.10 Irudia).



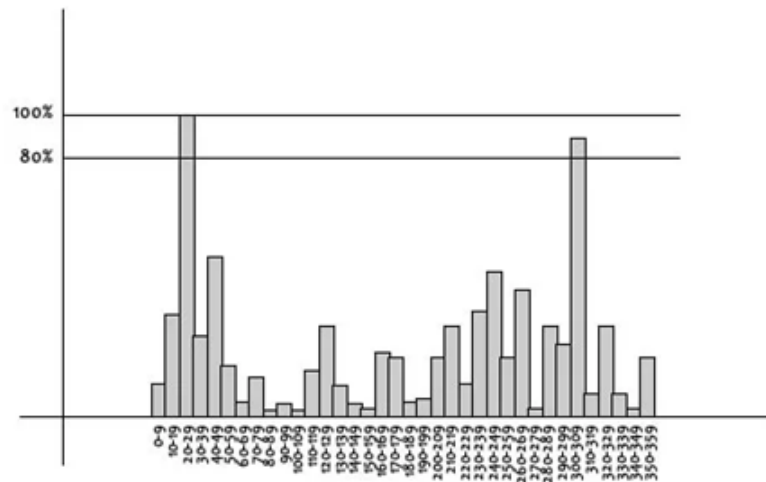
3.10 Irudia: X-z markatutako pixela berdez markatutako pixelekin alderatzen da leku potentziala den ala ez zehazteko.

Giltza-puntuak zehaztea

Behin puntu potentzialak lortuta, horietatik giltza-puntuak zein izango diren zehazten da, hau da, irudiaren ertzean daudenak eta kontraste nahikorik ez dutenak baztertu behar dira. Kontraste nahikorik ez dutenak identifikatzeko, Taylorren serieak erabiltzen dira, erpinaren maximo edo minimo bakoitzaren intentsitatea zehaztu, eta, muga batetik behera badago (0,03 txosten originalean [Lowe, 1999]), baztertua izango da. Aldiz, matrize hesiarrak erabiltzen dira irudiaren ertzean agertu zaizkigun puntu potentzialak bahetzeko, LoG sentikorregia baita horietan.

Giltza-puntuei orientazioak esleitzea

Giltza-puntu bakoitzari orientazio bat esleitzen zaio, biraketekiko inbariantza izateko. HOGen bezala lortzen dira gradienteak, eta hortik histogramak (360° -koak kasu honetan beti, ez 180° -koak). Bada, histogramak punturen batean tontor bat izango du gutxienez, eta altuera maximoaren %80tik gorako tontor denak hartuko dira orientazioa kalkulatzeko (3.11 Irudia).



3.11 Irudia: Maximoaren %80 baino handiago diren balioak soilik hartzen dira kontuan.

Giltza-puntuak bektore multidimentsionalez deskribatzea

Puntu honetara iritsita, HOGeko pauso berdin-berdinak jarraitzen dira: gelaxkatan bantzea, histogramak lortzea eta abar, baina giltza-puntuaren inguruetan soilik, ez irudi guztian. Gainera, biraketarekiko dependentzia kentzeko, gradiente bakoitzari giltza-puntuaren errotazioa kentzen zaio. Hartara, giltza-puntu bakoitzarentzat, bektore-deskriptore bat egongo da.

Giltza-puntuen arteko elkartzea

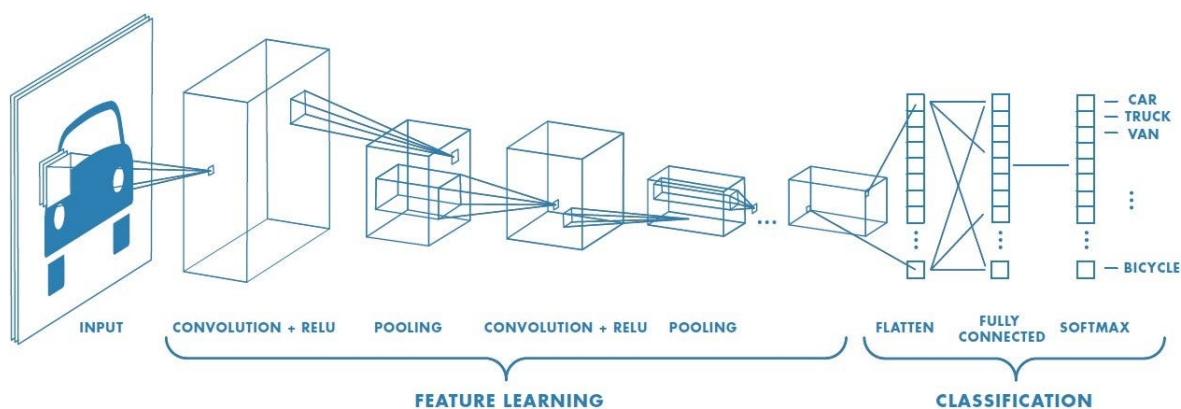
Bi irudien arteko giltza-puntu berdinaren elkarketa egiteko gertuen daudenak erabiltzen dira, baina batzuetan, bi oso gertu egon daitezke. Horregatik, gertuen dagoen giltza-puntuaren eta bigarren gertuenaren arteko proportzioa ikusten da, eta, 0,8 baino handiagoa bada, baztertu egiten da. Modu honetan, elkarketa okerren %90 saihesten dira, eta zuzen dauden %5 soilik, txostenaren [Lowe, 1999] arabera.

3.2.3 Neurona-sare konboluzionalak (Convolutional Neural Networks edo CNN)

Neurona-sare konboluzionalak neurona-sare artifizial mota bat dira, zeinetan ezkutuko geruzak konboluzionalak diren (3.12 Irudia). Bestelako neurona-sare artifizialek bezala, CNN batzuek sailkapena dute helburu amaieran, alegia, irudietatik ezaugarriak erauzten dituzte, eta ezaugarri horiek erabiltzen dituzte irudia sailkatzeko. Horregatik, ezaugarrie-

tatik sailkapena egiteko zatian azalduko da metodo honen bigarren zati hori. Zati hauek izan ditzake neurona-sare konboluzional batek:

1. Sarrera-irudia.
2. Ezaugarri-ikasketa.
 - a. Geruza konboluzionalak.
 - b. Pooling edo informazio-murrizketak.
3. Sailkapena.



3.12 Irudia: Neurona-sare konboluzionalen egitura. [Saturn Cloud, 2018]

Lehenengo geruzak pixeletako informazioa jasoko du, hain zuzen, pixel bat sarrera-neurona bat izango da. Ondoren, ezkutuko geruzek filtro asko izan ditzakete, eta filtro horiek matrizeak izaten dira. Demagun 3x3ko matrize bat hartzen dela filtro gisa, bada, irudian 3x3ko bloke bakoitzarekin Frobenius biderketa eskalarra egin, zenbaki berria gorde eta blokea desplazatuko da. Azkenean, zenbaki desberdinez osatutako beste matrize bat lortuko da (3.13 Irudia).

| | | | | |
|---|---|-----------------|-----------------|-----------------|
| 1 | 1 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 1 _{x1} | 1 _{x0} | 1 _{x1} |
| 0 | 0 | 1 _{x0} | 1 _{x1} | 0 _{x0} |
| 0 | 1 | 1 _{x1} | 0 _{x0} | 0 _{x1} |

| | | |
|---|---|---|
| 4 | 3 | 4 |
| 2 | 4 | 3 |
| 2 | 3 | 4 |

3.13 Irudia: Ezkutuko geruzetan filtroak aplikatzea matrize osoa zeharkatuz. [DimitrisMi, 2023]

Desbideratze bat eta aktibazio-funtzio ez-lineal bat ere aplikatu ohi zaizkio, beraz, orokortuz eta matematikoki idatziz, honela kalkulatzen da neurona konboluzional bakoitzaren irteera:

$$Y_j = g \left(b_j + \sum_i K_{ij} \otimes Y_i \right)$$

non Y_j j neuronaren irteera-matrizea den, zeina Y_i -ren (aurreko geruzako neuronen irteeraren) eta K_{ij} -ren (lotura bakoitzari dagokion pisuaren, hau da, filtroaren) arteko konbinazio linealarekin kalkulatzen den. b_j desbideratzea gehitzen zaio, eta, ondoren, g aktibazio-funtzio ez-lineal batetik pasatzen da.

Esaterako, 3.13 Irudian 3.14 Irudiko matrizeak daude.

$$Y \text{ (aurreko geruza)} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$K \text{ (filtroa)} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$Y \text{ (hurrengo geruza)} = \begin{bmatrix} 4 & 3 & 4 \\ 2 & 4 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

3.14 Irudia: Ezkutuko geruzetan filtroak aplikatzean erabiltzen eta lortzen diren matrizeak.

Hori jakinda, lehenengo ilaran kalkulu hauekin lortu dira geruza berriko balio berriak (desbideratzerik eta aktibazio-funtziorik gabe):

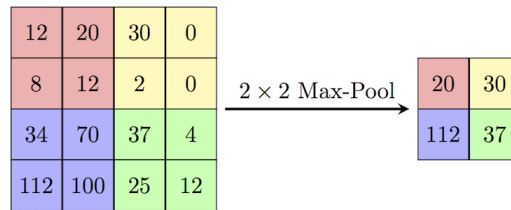
$$4 = 1 * 1 + 1 * 0 + 1 * 1 + 0 * 0 + 1 * 1 + 1 * 0 + 0 * 1 + 0 * 0 + 1 * 1$$

$$3 = 1 * 1 + 1 * 0 + 0 * 1 + 1 * 0 + 1 * 1 + 1 * 0 + 0 * 1 + 1 * 0 + 1 * 1$$

$$4 = 1 * 1 + 0 * 0 + 0 * 1 + 1 * 0 + 1 * 1 + 0 * 0 + 1 * 1 + 1 * 0 + 1 * 1$$

Sarea entrenatuz, azkenean, irudi arruntetako datu esanguratsuak informazio-matrize berrietan nabarmenago bihurtzen dira, eta sailkapenerako informazio lagungarria lortzen da.

Informazio esanguratsua lortzeko fasearen parte da *pooling* edo informazio-murrizketa ere. Azken finean, informazio-murrizketarekin bi irudi ia berdinetatik informazio esanguratsuen hartzean, informazio bera erazte lortzen da. Gaur egun, adibidez, *max-pooling* metodo eraginkorra erabiltzen da. *Max-poolingek* filtroetatik lortu ditugun informazio-matrize horietako zatiak hartu, eta beraien arteko zenbaki handienarekin ordezkatzan dituzten matrize txikiagoetan, informazioa murriztuz (3.15 Irudia).



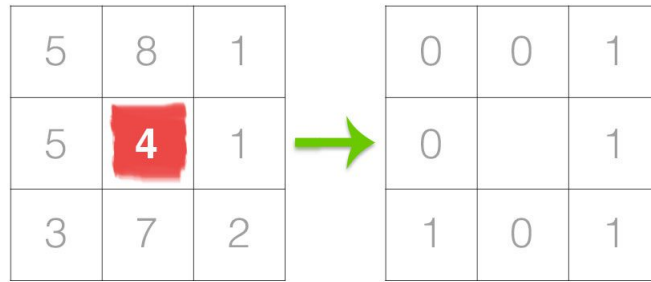
3.15 Irudia: *Max-pooling*en funtzionamendua.

Behin informazio-erazketa fase bat edo gehiagoren ostean, sailkapen-faseari ekin dakioke.

3.2.4 Eredu Bitar Lokalak (*Local Binary Patterns* edo LBP)

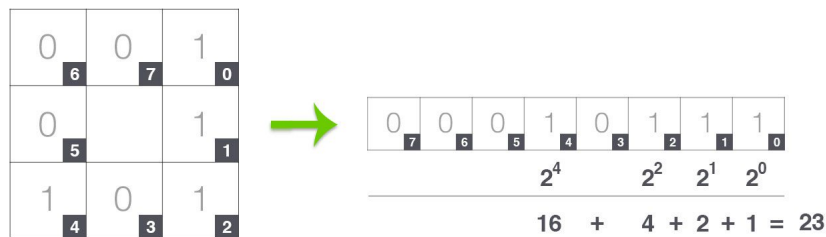
Irudietatik ezaugarriak ateratzeko beste metodo bat LBP gisa ezagutzen dena da, zeina lehenengoz 1994. urtean definitu zen [Ojala et al., 1994]. Gainera, arestian aipaturiko HOG metodoarekin konbinatuz, datu-base batzuetan oso ongi funtzionatu duela ikusi da. Bertsio desberdinak dituen metodoa da, baina txosten oso batean jasota daude hainbat mota [Bouwman et al., 2018].

LBP mota ohikoenak pixel bakoitza bere inguruko 8 pixelekin alderatzen du. Erdiko pixelaren balioa handiago denean, 0 bat ezartzen du, eta, bestela, 1 bat, zortzi horietako bakoitzean (3.16 Irudia), horrela, 8ko luzera duen bektore bitar bat lortuz.



3.16 Irudia: LBP adibide baten lehen pausoa, inguruko zenbakiekin alderatzea.

Bektorea lortzeko zenbakiak goi-eskubiko baliotik hartzen hasi eta erlojuaren noranzkoa jarraitzen da. Ondoren, bektore bitarra zenbaki hamartar bihurtzen da (3.17 Irudia).



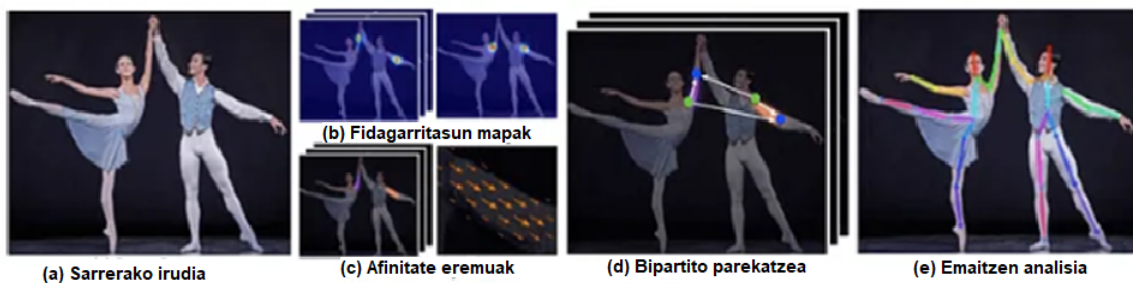
3.17 Irudia: Matrizetik bektorea eta bektoretik zenbaki hamartarra lortzea.

Horrenbestez, hasierako pixelari zegokion leku berean dagokion zenbaki hamartarra jarriko da, eta, behin zenbaki hamartar guztiak lortuta, 0tik 255era arteko zenbakiak izango direnez, 256 luzerako bektore karakteristiko batean jaso dezakegu zenbaki bakoitza zein maiztasunekin agertzen den irudian. Hartara, irudia bektore karakteristiko batean jasotzea lortzen da.

3.2.5 Eskeletoak

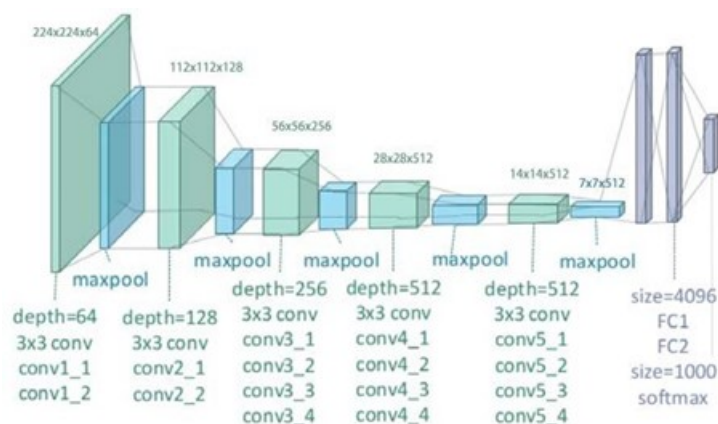
Histogramak edota bektore karakteristikoak ohikoak diren arren, giza ekintzak identifikatzeko garaian, eskeletoak lortzeko metodoak erabiltzen dira. Horiek lortzeko metodo desberdinak daude, aurrez ikusitako CNN eta giltza-puntuak lortzeko metodoak erabiltzen dituzte gaur egun erabiliak diren aplikazio askok. Horregatik, jada metodo garrantzitsuenak azalduta daudenez, hauen konbinazio bat azalduko da irudietatik eskeletoetara iristeko. Gainera, lortzen diren eskeletoen inguruko azalpenak emango dira, giza ekintzen identifikazioan ezinbestekoa den informazio-mota baita.

Modeloen prozesua aztertu aurretik ulertu behar da eskeletoak determinatzeko bi ezaugarri nagusi hartzen direla kontuan: fidagarritasun-mapak (*Part Confidence Maps*) eta afinitate eremuak (*Part Affinity Fields*) (3.18 Irudia). Fidagarritasun-mapak zehazten dute pixel jakin batek duen probabilitatea giza gorputzeko zati bat izan dadin, eta afinitate-eremuak giza gorputzaren orientazioa zehazten duten bektore-eremuak dira. Bestalde, identifikazioen zailtasuna oso desberdina da irudian eskeleto bakarra dagoenean edo gehiago dagoenean. Orain, OpenPose [Cao et al., 2019] [Cao et al., 2017] [Simon et al., 2017] [Wei et al., 2016] arkitekturaren funtzionamendua azalduko da, batik bat jende askoren eskeletoak identifikatzea nola ahalbidetzen duen ikusteko, eta, bestetik, oso erabilia delako.



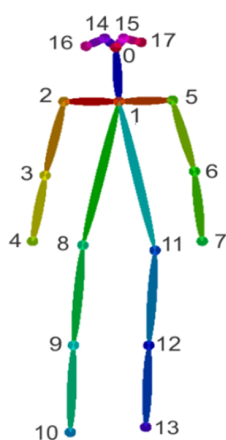
3.18 Irudia: OpenPose arkitekturaren *pipeline*. [Cao et al., 2017]

Lehenik eta behin, irudia VGG-19 ereduaren [Simonyan and Zisserman, 2015] lehen 10 geruzatetik igarotzen da. VGG-19 ereduaren 19 geruzak dituen neurona-sare konboluzionala da (3.19 Irudia), eta ezaugarri mapa desberdinak lortzen dira.



3.19 Irudia: VGG-19 eredu osoa. [Zheng et al., 2018]

Ondoren, ezaugarri mapa horiek bestelako CNN batzuetan prozesatzen dira, eta aurrez aipaturiko konfiantza-mapak eta afinitate-eremuak lortzen dira. Artikulazio bakoitzak bere konfiantza-mapa du, eta honako hemezortzi artikulazioak bereizten dira (3.20 Irudia): sudurra (0), lepoa (1), eskuin eta ezker sorbaldak (2, 5), eskuin eta ezker ukondoak (3, 6), eskuin eta ezker eskumuturrak (4, 7), gerriaren eskuin eta ezker aldeak (8, 11), eskuin eta ezker belaunak (9, 12), eskuin eta ezker oinak (10, 13), eskuin eta ezker begiak (14, 15) eta eskuin eta ezker belarriak (16, 17).



3.20 Irudia: OpenPosek eskeletoan identifikatzen dituen artikulazioak.

Azkenik, algoritmo jale (*greedy algorithm*) batean prozesatzen dira konfiantza-mapa eta afinitate-eremuak, irudiko pertsona bakoitzarentzat gorputz-jarrerak lortzeko. Hala eta guztiz ere, zehaztuko da nola desberdintzen diren pertsona desberdinen eskeletoak. Konfiantza-mapetatik artikulazio guztiak aurkitzen dira, eta, afinitate-eremuetatik abiatuta, artikulazio haueetatik zein doazen batera zehazten da. Bigarren pauso horretarako, artikulazioen arteko konexio bakoitzaren jatorrizko artikulazioa, helmugakoa eta konexioa ona den baloratzen puntuazio bat gordetzen dira, eta, horren arabera, gorputz-adarrak lortu. Azkenik, informazio hori guztia erabilita lortzen dira eskeleto osoak.

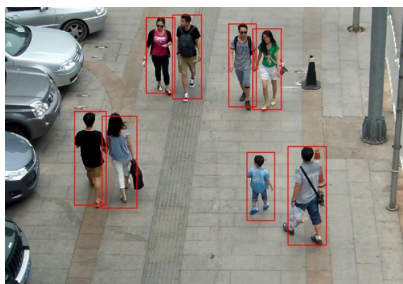
Aurrerago ikusiko denez, landuko ditugun irudietan aparkaleku batean gertatzen diren ekintzak ikusiko dira, hori dela eta, aipatzekoa da azaldu den algoritmoa ibilgailuetan ere erabiltzen hasi dela, alegia, errepideko ibilgailuen eskeletoak lortzeko erabiltzen dela (3.21 Irudia).



3.21 Irudia: Ibilgailuetan giltza-puntuen detekzio adibidea. [Cao et al., 2017]

3.2.6 Kaxa inguratzaileak (*Bounding Box* edo *BBox*)

Objektu edota gizakiek irudietan duten kokapena zehazteko eskeletoak baino egitura sinpleagoak dira *bounding boxak* (3.22 Irudia). *Bounding box* batek irudi batean objektu edo pertsona baten mugak zehazten ditu; alegia, objektu edo pertsona kaxa batez inguratzen du. Lau balio nagusik zehazten dute kaxa normalean, zehazki, kaxaren goi-ekerreko eta behe-eskuineko puntuen koordenatuek. Halere, batzuetan, goi-ekerreko puntuen koordenatuak, kaxaren zabalera eta kaxaren altuerarekin definitu ohi dira.



3.22 Irudia: *Bounding box* identifikazio-adibidea.

Informazio-egitura sinpleak izan arren, pertsonen posizioa, zabalera eta altuera informazio esanguratsua izan daiteke, eta, horrez gain, desplazamendu-abiadura ere atera daiteke irudi-segida batean.

3.3 Sailkapen gainbegiraturako teknikak

Behin irudietako ezaugarriak datu-egituretan daudela, horiek interpretatzeko metodoak behar dira sailkapena gauzatzeko. Esan bezala, gure proiektuan, *bounding boxak*, eskeletoak eta hauetatik erauzitako atributuak erabili dira sailkatzaileei pasatzeko, hau da, ez da bektore karakteristiko edo histogramarik lortu (nahiz eta eskeletoak lortzeko prozesuan erabili diren). Lan honetan, *Multilayer Perceptron* [Haykin, 1994] neurona-sare bat erabili da, eta, beraz, horien azalpenarekin hasiko gara.

3.3.1 *Multilayer Perceptron* neurona-sareak (MLP)

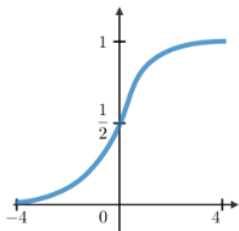
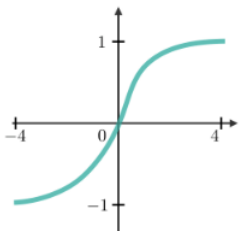
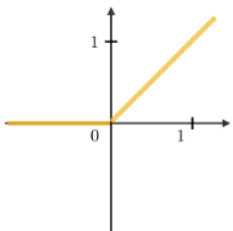
Neurona-sare mota hau eta aurrez ikusitako konboluzionala desberdinak dira. Konboluzionalek ez bezala, aurreko geruzako neurona guztiak hurrengo geruzako neurona guztiarekin elkartuta daude (*fully connected*). Ondorioz, parametro gehiago erabiltzen dituzte, eta, batzuetan, erredundantea izan daiteke. Irudien sailkapenean, konboluzionalak erabiliagoak dira, baina MMPose eta eskeletoekin lan egitea erabaki denez, guk eraikitako sareak jasotako datuak ez dira bere horretan irudiak izango, kasu honetan (MMPosek bai irudiak jasoko dituela, baina eraikiko den neurona-sareak eskeletoak soilik). Hori dela eta, hautatu da *multilayer perceptron* motako neurona sare bat erabiltzea eskeletoak lortu ondoren.

Izenak dioenez, MLP motako neurona-sareak elkarri lotutako hainbat neurona geruzaz osatutako sareak dira. Geruza desberdinetako neuronen arteko lotura bakoitzean, pisuak daude, kalkuluertarako erabiltzen direnak. Aipatutako geruza horietan hiru mota bereizten dira: sarrera-geruza, ezkutuko geruzak eta irteera-geruza. Sarrera-geruzetan aldagaien balioak jasotzen dira; hortaz, aldagai bakoitzeko neurona bat behar da. Ezkutuko geruzetan, aurreko geruzatik jasotako balioekin eta pisuekin kalkuluak egin eta hurrengo geruzetara pasatzen dira balio berriak, harik eta irteera-geruzara iristen diren arte. Irteera-geruzan, normalean, neurona bakoitza klase bati dagokio, eta, klase bakoitzak jasotzen duen balioaren arabera, hasierako instantzia klase batera edo bestera sailkatzen da. Gerta daiteke, irteeran, klase batera sailkatzea lortu nahi izan ordez, balio konkretu bat lortu nahi izatea (erregresioa); horregatik, ez dago azken geruzan beti klase-kopurua, ez dagoelako emaitza kopuru finitu bat.

Sare hauek entrenatzean, egindako erroreen arabera, arestian aipaturiko loturetako pisuak aldatzen dira, errorea murriztuz pixkanaka, hau da, errore-funtzioa minimizatuz. Horretarako, ikasketa gainbegiraturako teknikak erabiltzen dira, benetako emaitza jakin ondoren eta lortutako errorearen arabera, atzekoz aurrerako algoritmo (*backpropagation*) baten bidez pisuak eguneratuz. Algoritmo desberdinak daude errore-funtzio hau minimizatzeke. Ezagunenak gradiente-jaitsiera [Ruder, 2016], Adam [Kingma and Ba, 2014] eta LBFGS [Liu and Nocedal, 1989] dira. Gradiente-jaitsierak pisuak aldatzen ditu errore-funtzioaren balio baxuenak non dauden ondorioztatuz iteratiboki; gradiente-jaitsiera desberdinak daude, eta erabilienetako bat gradiente-jaitsiera estokastikoa da. Adam (*Adaptive Moment Estimation*) algoritmoak gradiente-jaitsieraren antza du, baina konbergentzia azkarragoa du, bestelako aldagai batzuk inplementatzen dituelako iterazio bakoitzean. Azkenik, LBFGS algoritmoa matrize hessianez baliatzen da, funtzioaren konbergentzia horretako pausoen

noranzkoa eta tamaina definitzeko.

MLP bat eraikitzean definitu beharreko beste aldagai bat aktibazio-funtzioa da. Aktibazio-funtzioa funtzio ez-lineala da, geruza bakoitzaren irteeran aplikatzen dena, eta zenbakien arteko kalkuluak kalkulu lineal soilak izatea ekiditen da. Aktibazio-funtzioen mota anitz dago, ezagunenak sigmoide funtzioa, tangente hiperbolikoa (tanh) eta ReLU funtzioak dira (3.23 Irudia).

| Sigmoid | Tanh | RELU |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| $g(z) = \frac{1}{1 + e^{-z}}$ | $g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ | $g(z) = \max(0, z)$ |
|  |  |  |

3.23 Irudia: Aktibazio-funtzio ezagunen formulak eta grafikatzak.

Hiperparametroak ere definitu behar dira, esate baterako, egingo diren iterazio-kopuru maximoa edo pisuen hasieraketa; izan ere, hasierako pisuak ausaz jartzen badira exekuzio bakoitzean, sarrerako datuak berak izanda ere, emaitza desberdinak lortuko dira, eta egiten diren konparaketak ez dira baliagarriak izango. Behin sarea eraikita, hasieran ditugun datu guztietatik erabaki behar da ehuneko zenbat izango den entrenatzeko, (*train*) eta zenbat, probak egiteko (*test*).

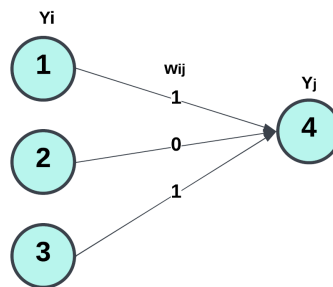
3.3.2 Neurona-sare konboluzionalak (*Convolutional Neural Networks* edo CNN)

Aurrez esan bezala, neurona-sare konboluzionalen sailkapeneko zatia orain azalduko da. Gogora dezagun ezaugarriak lortzeko fase bat edo gehiagoren ostean datorren fasea dela, hau da, hasierako irudietatik jada datu-egiturak lortu dira. Sailkapeneko zatia lortutako datuei etiketa bat edo beste jartzea erabakitzean datza, eta zati horretarako neuronak aurrekoek bezala funtzionatzen dute:

$$Y_j = g\left(b_j + \sum_i w_{ij} \cdot Y_i\right)$$

non j neuronaren Y_j irteera aurreko geruzako Y_i eta beraien arteko w_{ij} pisuen arteko konbinazio lineala den. b_j desbideratzea gehitzen zaio, eta g aktibazio-funtzio ez-lineal batetik pasatzen da. Demagun aurreko geruzan hiru neurona ditugula, 3.24 Irudian ikus daitezkeenak: 1, 2 eta 3 balioekin eta hurrengo geruzako neurona batera elkartzen dituen loturek 1, 0 eta 1 pisuak dituzte hurrenez hurren. Bada, honela kalkulatu litzateke hurrengo geruzako neurona horren pisu berria (desbideratzerik eta aktibazio-funtziorik gabe):

$$1 * 1 + 2 * 0 + 3 * 1 = 4$$



3.24 Irudia: Adibiderako sortutako neurona-sarearen irudikapena.

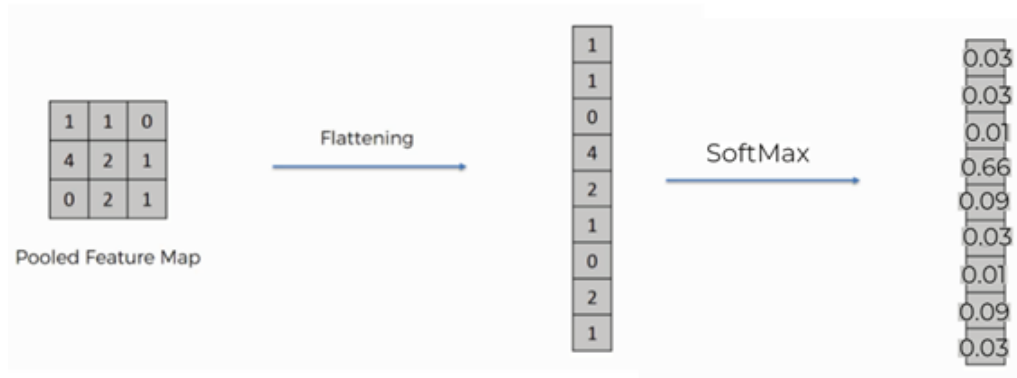
Sailkapen-zatian ere are gehiago aldatzen dira datuak. Horretarako, lautzea edo *flattening* izeneko prozesuan bektore moduan jartzen dira matrize moduko datuak. Bestalde, batzuetan, bektore hauetako neurona bakoitza hurrengo geruzako guztiekin lotzen da, eta, pisu ezberdinen arabera, zenbaki berri desberdinak lortzen dira geruza berri horietan. *Fully connected* izena dute neurona guztiak lotuta dauden geruza-segida hauek. Amaitzeko, softmax eta antzeko funtzioak oso ohikoak dira, probabilitate bat lortzen baita klase bakoitzerako. Ondoren, probabilitate handiena duen klasea esleitzen zaio hasieran jasotako irudiari; hitz gutxitan, sailkatu egiten da. Honela definitzen da softmax matematikoki:

$$\sigma : \mathbb{R}^D \rightarrow [0, 1]^D$$

$$\sigma(z)_j : \frac{e^{z_j}}{\sum_{d=1}^D e^{z_d}} \quad j = 1, \dots, D.$$

Hori aplikatuta, zenbaki errealak dituen D dimentsioko bektore bat dimentsio bereko bektore bihurtzen da, baina balioak $[0, 1]$ tartekoak dituenak. Balio berri guztien baturak 1

ematen du, eta, beraz, probabilitate-banaketa jakin bat lortzen da bektoreko balioen artean. Esan bezala, probabilitate handieneko klasera sailkatuko da sarrera; adibidez, [3.25](#) Irudian, 4. klaseko instantzia gisa sailkatuko litzateke sarrera (0,66 baita balio handiena).



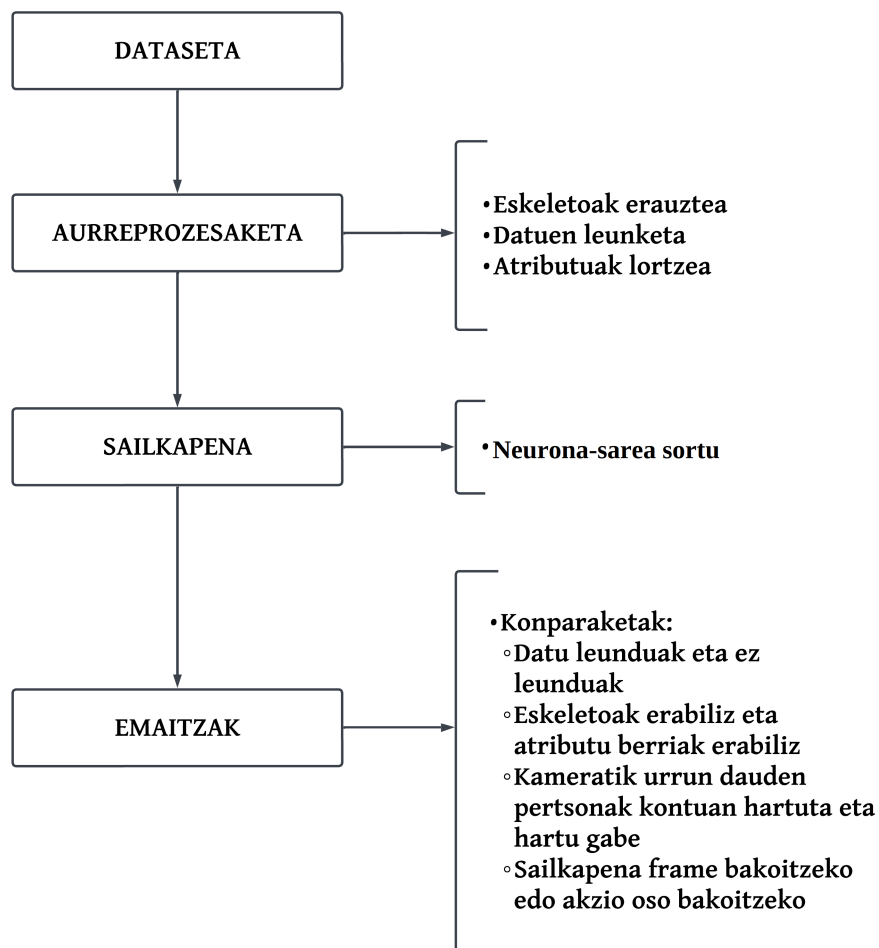
3.25 Irudia: Flattening eta Softmax erabilia egindako sailkapen bat.

4. KAPITULUA

Egindako esperimentazioa eta lortutako emaitzak

Behin teknika sorta bat ikusita, lan honetan erabilitakoak zehaztu daitezke, eta hasierako datuetatik amaierako sailkapenera nola iritsi den azaltzen duen prozesua azaldu daiteke. Beraz, datu-multzoa deskribatzetik hasi eta lortutako emaitzetara arte iritsiko da azterketa (4.1 Irudia).

PROIEKTUAREN ESKEMA OROKORRA



4.1 Irudia: Proiektuaren eskema orokorra edo pipelinea.

Hain zuzen ere, erabili den dataseta nolakoa den eta nondik lortu den azalduko da. Ondoren, datu horiek guztiak zein aurreprozesaketa-fasetatik pasa diren ikusiko da, hau da, eskeletoen erauzketa, datuen leunketa eta atributuen erauzketa azalduko dira. Horren ostean, eraikitako neurona-sarea deskribatuko da, eta, azkenik, arkitektura hau erabiliz egindako proba desberdinak eta lortutako emaitzak ikusiko dira.

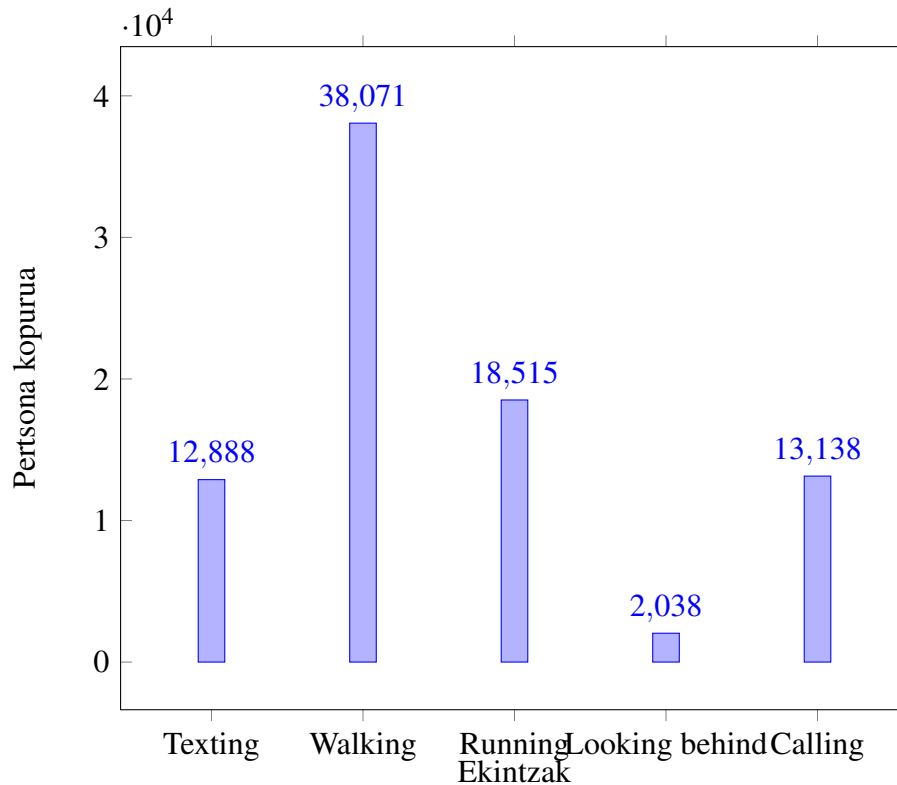
4.1 Erabilitako datu-multzoa

Jasotako datu-basea Irlandako Galwayeko Unibertsitateko Connaught Automotive Research Group ikerketa-taldeak [Connaught Automotive Research (CAR) Group, 2005] utzitakoa izan da. Ekintza-errekonozimenduaren inguruko proiektu pribatu batean erabiltzen ari diren datu-basea da. Zehazki, aparkaleku bateko bideo bat jaso da, segurtasun-kamera baten ikuspuntutik grabatuta, eta hainbat pertsona ekintza desberdinak egiten agertzen dira. Bideo horiek konfidentzialak izanik, lan honetan ezin izango dira bertako irudiak erabili, baina antzeko irudiak erabiliko dira testuingurua ulertarazteko.

Datu-basea bere horretan bideo bat izanda ere, arestian aipaturiko tekniketari azaldu bezala, bestelako informazio bat atera behar izan da bideoetatik. Kasu honetan, *bounding box* eta eskeletoak landu dira, ondoren sailkapen prozesuari ekin ahal izateko. Horregatik, bideoa oinarri izan arren, bideoan zehar ekintzak egiten dabiltzan pertsonen *bounding box*ak eta egiten ari diren ekintzak etiketatuta jaso dira. *Bounding box*ak goi-ezkerreko puntuaren koordenatuekin eta zabalera eta altuerarekin definituta jaso dira. Ekintzen etiketei dagokienez, bost ekintza desberdin sailkatu nahi izan dira: pertsona oinez, korrika, atzera begira, telefonoz hitz egiten eta mugikorrean mezu bat bidaltzen. Azken finean, aparkaleku baten testuinguruan, istripu edota arrisku egoerekin lotuta egon daitezkeen ekintzak dira, arreta beste zerbaitetan izatea eskatzen dute eta. Beraz, hitz gutxitan, bideo bat da, bost ekintza horiek egiten ari diren pertsonekin eta beraien *bounding box*ekin eta ekintza etiketatuekin.

Zehazki, 84.650 instantzia daude, eta bakoitzak jasotzen ditu pertsona baten kaxa ingurutzaila, agertzen den *frame*aren identifikatzailea (id), pertsonaren id-a eta egiten ari den ekintzaren etiketa. Ekintza bakoitzeko instantziak zenbatu ondoren, honakoak dira instantzia kopuruak: 12.888 pertsona mugikorrean mezu bat bidaltzen (*person texting*), 38.071 pertsona oinez (*person walking*), 18.515 pertsona korrika (*person running*), 2.038 pertsona atzera begira (*person looking behind*) eta 13.138 pertsona telefonoz hitz egiten (*person on a call*). Aurreprozesaketa egin aurretik, esan dezakegu datuak desorekatuta daudela eta

4.2 Irudian ondo ikusi daiteke hori.



4.2 Irudia: Ekintza bakoitzaren instantzia-kopurua datu-basean.

Instantzien *frame_id*-ei dagokienez, 20tik 10.799ra arteko balioak hartzen dituzte (20. *frame*-ean hasten baitira ekintzak eta ez lehenago). *Person_id*-ek 28tik 310era arteko balioak hartzen dituzte, eta esan beharra dago, bideoako pertsona berdin batek ekintzaz aldatzen duenean, *person_id*-a aldatzen zaiola. Bada, behin datuak ulertuta, aurreprozesaketarekin hasi gaitzke.

4.2 Aurreprozesaketa

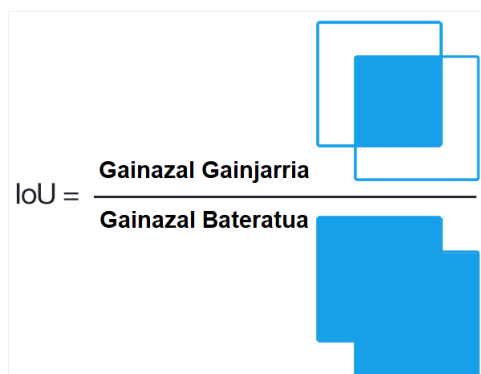
4.2.1 Eskeletoak erauztea

Jasotako datuak edo instantziak prozesatu ahal izateko, lehenengo, instantzia akastunak ezabatu egin behar dira, behar ez direnak kendu eta antzeko baheketak egin. Bideoak eta datuak aztertu ostean, aurreprozesaketa egin behar izan da eskeletoak lortzeko prozesuetan eta lortu ostekoetan, hau da, sailkapenaren aurreko prozesuetan. Horregatik, atal honetan ikusiko dira bai eskeletoak lortzeko prozesua eta baita lortu ondoren egindako baheketa ere.

Hasteko, azalduko da nola lortu diren bideoko pertsonen eskeletoak. Horretarako, MMPose [Cao et al., 2018] erabili da, PyTorch softwarean oinarritzen den erreminta-kutxa (*toolbox*). Zehazki, eskeleto bakoitzarentzat, 17 giltza-puntu ateratzen dituen COCO eredua erabili da. Software berak bere *bounding box* propioak ere kalkulatu ditu, eta *bounding box* hauek erabili dira etiketak dituzten instantziei MMPosek lortutako eskeletoak esleitzeko, jarraian azalduko den bezala.

Bi datu base zeuden: Galwayeko Unibertsitatetik eskuz lortutako hasierako datu-basea (pertsona bakoitzeko bere *bounding box*arekin eta etiketarekin) eta MMPose bidez lortutako datu-basea (pertsona bakoitzeko MMPosek lortutako *bounding box*a eta eskeletoarekin). Beraz, eskuz lortutakoek ez dute eskeletorik, horregatik, MMPose bidez lortutakoak esleitu nahi izan zaizkie. Horretarako, nahikoa litzateke datu-base bakoitzeko *bounding box*ak aztertzea eta balio berdinak dituztenekin bikoteak egitea, horrela, bi datu-baseak bateratuko genituzke instantziaz instantzia, eta eskuz etiketatutako pertsona bakoitzak bere MMPose eskeletoa izango luke.

Halere, eskuz lortutako *bounding box*ak eta MMPose bidez lortutakoak ez dira berdindin balioei dagokienez, horregatik, ezin dira balio berdinak zuzenean bilatu eta elkartu. Esan bezala, egiazkotzat edo ontzat (*ground truth*) hartu diren *bounding box*ak etiketadunak izan dira, beraz, hauetako bakoitza banan-banan hartu eta MMPoserekin lortutako guztiekin konparatu da egokia esleitzeko. Esleitu behar zaiona zein den definitzeko, hainbat modu daude, baina *Intersection over Union* edo IoU teknika erabili da (4.3 Irudia). Teknika horrek bi *bounding box* beraien artean zenbateraino gainjartzen diren zehazten du. Horretarako, gainjartzen den gainazala eta bi kaxek estaltzen duten gainazal osoaren arteko zatiketa egiten du; izan ere, zenbat eta berdinago izan bi gainazal horiek, zatiketaren emaitza orduan eta gertuago egongo da 1etik.



4.3 Irudia: IoU kalkulua funtzionamendua. [Rosebrock, 2016]

Horrela, eskuzko datu-baseko instantzia bakoitzari, *frame* berean baina beste datu-basean IoU balio handiena duen instantzia esleitzen zaio, hau da, antzekoena dena. *Bounding box* bikote batzuk oso gutxi gainjartzen dira, hau da, zaila da esatea benetan kaxa hori zegokion. Kasu horiek ekiditeko, IoU-ren balioa maximoa 1 dela kontuan hartuta, gutxienez 0,3ko IoU balioa lortu beharra ezarri da, alegia, bikoteak gutxienez %30ean gainjartze behar dira, bestela bikoterik gabe utzi dira. Beraz, balio minimo horregatik eta beste arrazoi batzuegatik (MMPosek eskeletoren bat ongi ez detektatzea, adibidez), instantzia batzuk bikoterik esleitu gabe geratu dira eta datu-basetik ezabatu dira. Esan beharra dago, bateratze hau egin aurretik, MMPosek emandako datu-basetik, hasieran, instantzia batzuk kendu behar izan zirela, eskuz lortutako datu-basean hasieratik ez baitzeuden instantzia horiek, 1etik 20rako *frame*ak kasu. Gainera, hasierako 84.650 instantzietatik 13.178k ez dute bikoterik. Aipatzekoa da, 0,3ko IoU minimoa jarri gabe, 12.058 zirela bikoterik gabeak. Ondorioz, hasierako 84.650 instantzietatik 71.472ra murriztu da datu-basea.

Aipatu behar da, halaber, lortutako eskeletoak ez direla ezaugarriak lortzeko erabiliko soilik; alegia, eskeletoen puntuetako posizioak bere horretan ere erabiliko dira neurona-sarean.

Bestela esanda, neurona-sarea bi atributu multzo desberdinekin entrenatuko da, bakoitza bere aldetik: batetik, eskeletoak definitzen dituzten 17 puntuen koordinatuekin soilik, eta, bestetik, eskeletoetatik erauziko diren bestelako atributuekin. Hain zuzen ere, neurona-sareek hasierako geruzan aldagai bakoitzeko neurona bat behar izaten dute, hau da, balio bakoitzeko neurona bat. Orduan, lehenengo sareak 34 neurona izango ditu lehenengo geruzan (eskeletoen 17 puntuetako bakoitzaren x eta y koordinatuak). Aldiz, erauziko diren atributuak 11 besterik ez direnez (aurreprozesaketa atalean (4.2) zehaztuko dira), 11 neurona soilik behar dira lehen geruzan. Horregatik, atributuetatik emaitza berak edo hobeak lortzen badira, ataza bera burutuko da, baina modu efizienteagoan, sare sinpleago batekin.

4.2.2 Eskeletoen leunketa

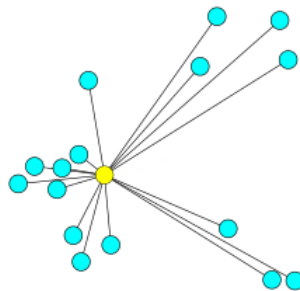
Lortutako eskeletoekin lan egin da, baina, bideoaren gainean eskeletoak irudikatu ostean, ikusi da MMPosek *frame* gutxi batzuetan eskeletoak errealtatetik oso urrun detektatzen dituela. Horregatik, puntuen kokapenak leundu nahi izan dira, hau da, aurretik eta ondoren dituzten kokapenaren arabera homogeneousatu.

Leunketa horretarako, lehenik eta behin, *frame*etako zenbat puntu hartuko diren kontuan definitu behar da, eta, ondoren, zein algoritmo matematiko aplikatuko den *frame* horietako puntuen leunketa gauzatzeko. Hamar *frame*etako leihoak definitzea erabaki da; alegia,

hamar *frameta*ko multzotan banatu da bideo osoa, eta, puntu bakoitzarentzat, leiho bereko gainontzeko bederatzi *frameta*n izan duen kokapena aztertu da. Algoritmo matematikoari dagokionez, mediana geometrikoa kalkulatzeko erabaki da (4.4 Irudia). Espazio euklidearrean kokatuta dauden puntu batzuen mediana geometrikoa puntu horietarako distantzien batura minimizatzen duen puntua da. Bestela esanda, mediana geometrikotik gainontzeko puntu guztietara dauden distantzien batura existitzen den txikiena da. Ondorioz, ez da puntuen batez bestekoaren bera, eta ezohiko balioak saihesteko aproposa da. Formalki, $\alpha_1, \dots, \alpha_n \geq 0$ pisuak dituzten $x_1, \dots, x_n \in \mathbb{R}^D$ puntuen mediana geometrikoa honela idatzi daiteke:

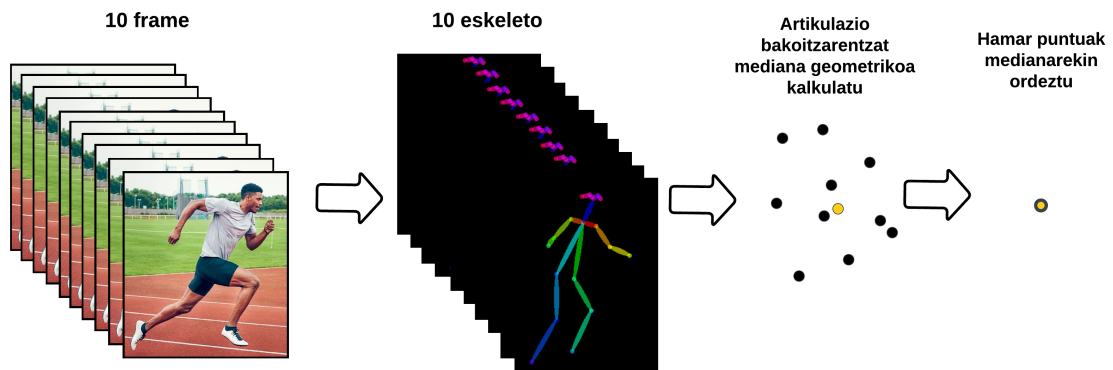
$$x_{MG} = \arg \min_{z \in \mathbb{R}^D} \sum_{i=1}^n \alpha_i \|z - x_i\|_2$$

Hemen, $\arg \min$ -ek batura minimizatzen duen y -ren balioa zehazten du. Kasu honetan, puntu bat da, zeinetatik gainontzeko x_i puntuetara dagoen distantzien batura minimoa den. Distantzia euklidearra norma gisa dago idatzita formulatan, eta 2-norma gisa ere definitzen da.



4.4 Irudia: Puntu-sorta baten mediana geometrikoa (horiz).

Ondorioz, hasierako eskeletoetatik eskeleto berriak lortu dira, puntuen kokapenak pixka bat eraldatuta eta leunduta (4.5 Irudia). Beraz, datu leundu eta ez-leunduaren arteko konparaketa egitea ahalbidetuko dute sailkapen-fasearen osteko datuek.



4.5 Irudia: Leunketa-prozesu oso baten adibidea.

4.2.3 Atributuak lortzea

Lehenengo garbiketa horren ostean, lortutako datu-basearen gainean lehen ezaugarriak lortzen hasi gaitezke. Bestela esanda, eskeletoak bere horretan erabili beharrean, esanguratsua izan daitekeen informazioa atera dezakegu bertatik. Informazioa esanguratsua izango den ala ez jakitea sailkapen-fasean ikusiko da, baina hipotesi batzuetatik abiatuta, potentzialki esanguratsuak diren ezaugarri batzuk lortuko dira. Betiere, kontuan izan behar da zein ekintza diren sailkatu beharrekoak, hau da, lortzen diren ezaugarriak ekintza zehatz horietan zeresana izatea hartu behar da aintzat.

Telefonoz hitz egitea

Hasteko, hipotesietako bat telefonoz hitz egiten dabiltzan pertsonak identifikatzeko egoia izan daiteke; zehazki, telefonoz hizketan gabiltzanean esku eta belarriaren arteko distantzia txikia dela ikus daiteke (4.6 Irudia). Hortik ondorioztatu dezakegu eskeletoaren belarri eta eskumuturraren arteko distantzia baliagarria izan daitekeela.



4.6 Irudia: Telefonoz hitz egitean belarri-eskuen arteko distantzia laburtzen da. Canva orrialdetik sortua [can, 2023], adimen artifiziala erabilia.

Distantzia hori kalkulatzeko, bi puntuen arteko distantzia euklidearra erabili da. MMPo-
seren eskeletoak erabili direnez, 3. puntua (ezker belarria), 9.a (ezker eskua), 4.a (eskuin
belarria) eta 11.a (eskuin eskua) erabili dira. Horrela, bi aldagai berri lortu dira: ezker
belarriaren eta eskuaren arteko distantzia, eta eskuin belarriaren eta eskuaren artekoa.

Sakelakoarekin mezuak bidaltzea

Beste ekintza batzuetatik ere hipotesiak ondorioztatu dira. Sakelakoarekin mezuak bi-
daltzen direnean, besoak tolesten dira mugikorra aurpegira hurbiltzeko, hau da, angeluak
ixten dira (4.7 Irudia). Hortik abiatuta, ondorioztatu da bi besoetako angeluak baliagarriak
izan daitezkeela sailkapenean laguntzeko.



4.7 Irudia: Telefonoz mezuak bidaltzean, besoetako angeluak ixten dira. Canva orrialdetik sortua [can, 2023], adimen artifiziala erabilia.

Kalkulu horietarako, beso bakoitzean bi bektore kalkulatu dira: ukondotik sorbaldara doana (ezker besoan, 7. puntutik 5.era; eta, eskuinean, 8. puntutik 6.era) eta ukondotik eskura doana (ezkerrean, 7. puntutik 9.era; eta, eskuinean, 8. puntutik 10.era). Bektore horiek lortzeko, bi puntuen koordinatuen arteko kenketak egin dira.

$$\vec{AB} = B - A$$

Ondoren, bi bektoreen arteko angelua honako formularekin lortu da:

$$\theta = \arccos \left(\frac{\vec{AB} \cdot \vec{BC}}{|\vec{AB}| * |\vec{BC}|} \right)$$

Kalkulu hauetan, arazo bat izan da, izan ere, bektore bat lortzerakoan bi puntu koordinatu berdinetan badaude, kenketak bektore nulua emango du, eta ezin izango da angelua kalkulatu beste bektore batekiko. Horregatik, kasu hauetan, instantzia osoa ezabatu da datu-basetik.

Korrika

Bestalde, antzeman da hankek eratzen dituzten angeluek ere zeresana izan dezaketela; esaterako, korrika gabiltzala, gehiago ixten direlako hankek eratzen dituzten angeluak (4.8 Irudia).



4.8 Irudia: Korrika goazenean, hanketako angeluak ixten dira. Canva orrialdetik sortua [can, 2023], adimen artifiziala erabilita.

Hanketako angeluetarako, besoetako kalkulu berak erabili dira, baina, kasu honetan, ezker hankarako, gerriko 11. puntua erabili da, belauneko 13.a eta oineko 15.a; eta, eskuinean, gerriko 12. puntua, belauneko 14.a eta oineko 16.a. Beraz, behin puntu honetara iritsita, *bounding box*aren goi ezkerreko puntuaren koordinatuak, *bounding box*aren altuera eta zabalera, beso-belarrien distantziak, besoetako angeluak eta hanketako angeluak lortu dira ezaugarri gisa.

Desplazamendu-abiadura

Azkenengo hipotesia desplazamenduarekin lotuta dago. Azken hipotesi honetan, pentsatu da teorian korrika doan pertsona bat oinez doan beste bat baino azkarrago doala. Horregatik, *frame* bakoitzarentzat, aurreko *frame*tik pertsona zenbat desplazatu den kalkulatu da. Pertsona bakoitzaren lehenengo *frame*ak ez duenez aurrekorik, 0 balioa esleitu zaie. Desplazamendua kalkulatzeko, distantzia euklidearra erabili da, *bounding box*aren (x, y) koordinatuetatik aurreko *frame*an pertsona berak zuen (x, y) koordinatuetara. Horrela, *pixel/frame* moduko abiadura bat lortu da. Halere, nahiz eta abiadura handiagoan ibili, kameratik urruti dagoen pertsona hurbil dagoena baino pixel gutxiago desplazatzen denez, pertsona bakoitzaren abiadura norberaren *bounding box*aren altuerarekin zatitu da, proportzionalagoa izan dadin.

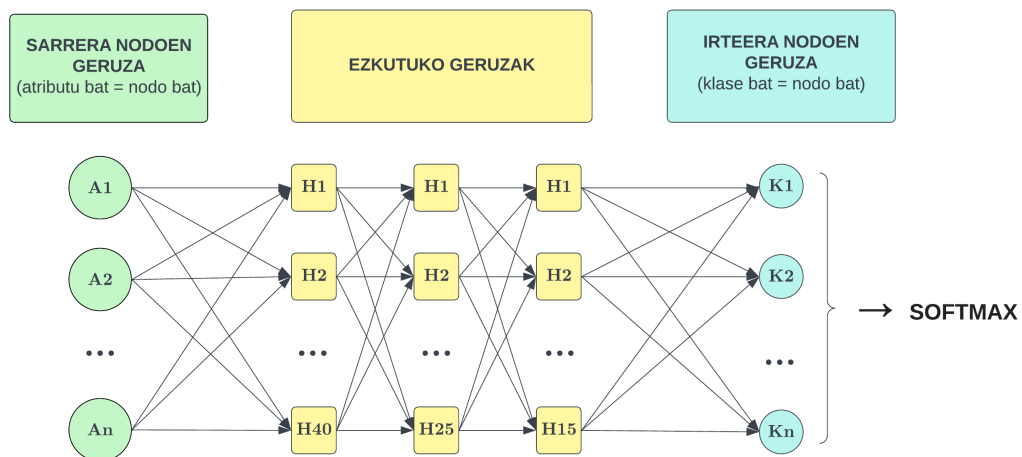
4.3 Sailkapena

Sailkapen-faseari ekiteko sailkatzaile bat hautatu behar da, eta neurona-sare bat eraikitzea erabaki da (4.9 Irudia). Atal honetan neurona-sarearen ezaugarri guztiak azalduko dira.

Hauek dira neurona-sarearen ezaugarriak:

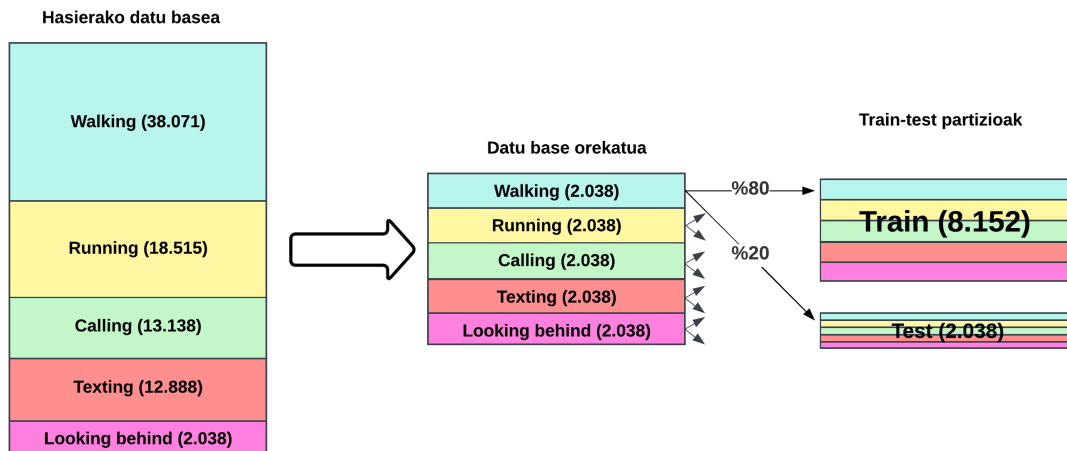
- Pythoneko scikit-learn liburutegiko MLPClassifierrekin eraikia [Pedregosa et al., 2011].
- *Seed* edo hazia (proba bera eginez gero, emaitza bera lortzeko): 31415.
- Sarrera-geruzan, atributu bakoitzeko sarrera-neurona bat.
- Hiru ezkutuko geruza: 40, 25 eta 10 neuronakoak, hurrenez hurren.
- Irteera-geruzan, 5 klase eta, beraz, 5 irteera-neurona.
- Iterazio-kopuru maximoa 1000 iterazio dira, konbergentzia lortu ahal izateko.
- Aktibazio-funtzioa: ReLU, tangente hiperbolikoak baino emaitza hobek eman ditu.

- Neuronetako pisuak esleitzen dituen algoritmoa: Adam, gradiente-jaitsiera estokastikoak baino emaitza hobek eman ditu. Milaka instantzietako datu-baseetan horrela izan ohi da.
- Amaieran, softmax funtzioa aplikatzen da.



4.9 Irudia: Eraikitako neurona-sarearen irudikapena.

Behin neurona-sarea eraikita, definitu behar da *train* eta *test* ataletarako nola banatuko den datu-basea. Esan bezala, lehenik datu-basea orekatu da, hau da, instantzia gutxien zituen ekintzak adina instantzia utzi dira beste ekintzetan ere, ausaz hautatuta. Ondoren, lehen probetan, datu-basearen %80 *trainerako*, eta gainontzeko %20a *testerako* utzi da, eta, partizio hauek sortzeko, instantziak ausaz banatzen dira. Ondorioz, erabaki da id bat zuen pertsona baten *frame* guztiak edo *trainen* edo *testen* osorik sartuko zirela, baina ez zirela bietan banatuko, horrela bermatzen baita antzeko kasuak ez direla bi partizioetan agertuko (4.10 Irudia).



4.10 Irudia: *Train* eta *test* partizioen banaketa.

4.4 Esperimentazioa

Instantziak sailkatzeko neurona-sarearen nondik norakoak jakinda, egindako probetan lortutako emaitzak, konparazioak eta hauen interpretazioak aztertzeari ekingo zaio. 4.1 Taulan jasotzen dira egindako probak eta lortutako asmatze-tasak.

Egindako proba guztiek logika bat jarraitzen dute. Hasteko, gogoratu behar da honako atributuak erauzi direla eskeletoetatik: *bounding box*ak, belarri eta eskuen arteko distantziak, besoetako angeluak, hanketako angeluak eta abiadura. Hori jakinda, lehenengo neurona-sarea atributu guztiekin entrenatu da, baina bakoitza bere aldetik erabilia, hau da, aldi bakoitzean atributu bakarria hartuta. Emaitza indibidual horietatik abiatuta, bi, hiru, lau eta bost (guztiak) onenak konbinatuz atera dira emaitzak, asmatze-tasaren bilakaera ikusteko.

Lehenago aipatu bezala, atributu guztiak konbinatzean lortutako asmatze-tasa eta eskeletoa bere horretan pasatzea konparatu dira, ea neurona-sarea, hasierako geruzan, neurona gutxiagorekin emaitza berak behintzat lortzeko gai den ikusteko. Bestalde, atributu guztiekin beste proba bat egin da, zeinetan hasierako datuak leunketa egin gabe pasatu diren.

Azkenik, hurbiltasun-probak egin dira, hau da, ikusi nahi izan da ea ereduak asmatze-tasa berdinak lortzen dituen kameratik hurbil dauden pertsonekin eta urruti daudenekin. Beraz, kameratik urrutien dauden pertsonen %25 eta %50 baztertu dira, emaitzak hobetzen ziren ikusteko, alegia, ea hurbileko pertsonak sailkatuta soilik hobeto lan egiten zuen aztertzeko. Gainera, hurbiltasun-proba hori berresteko, datuen %100 eta atributu guztiak

erabiltzean, asmatzen diren instantziak eta asmatzen ez direnak konparatu dira, zehazki, kameratik urruti dauden ala ez ikusteko. Horretarako, bi multzo horietako instantzien *bounding boxen* altuerak hartu, eta multzo bakoitzeko batez bestekoa kalkulatu da, hurbil daudenen *bounding boxak* handiago direlako, eta alderantziz.

4.5 Emaitzak

| SAREARI PASATAKO ATRIBUTUAK | ASMATZE-TASA |
|----------------------------------------------------------------|--------------|
| <i>Bounding Boxa</i> (x, y, altuera, zabalera) | 0,2 |
| Hanketako angeluak | 0,24 |
| Desplazamendu abiadura | 0,31 |
| Belarri/esku distantziak | 0,38 |
| Besoetako angeluak | 0,48 |
| B/E dist. + Bes. ang. | 0,49 |
| B/E dist. + Bes. ang. + abiadura | 0,59 |
| B/E dist. + Bes. ang. + abiadura + Hank. ang. | 0,59 |
| B/E dist. + Bes. ang. + abiadura + Hank. ang. + BBox (Guztiak) | 0,62 |
| Guztiak (datuak leundu gabe) | 0,61 |
| Guztiak (%75 hurbilenak) | 0,61 |
| Guztiak (%50 hurbilenak) | 0,68 |
| Eskeletoa bere horretan atributu gabe | 0,64 |

4.1 Taula: Esperimentazioetako datuak.

Aurrez esandako probak egin eta asmatze-tasak lortu ostean, interpretazioarekin hasi daiteke. Lehenik eta behin, ikus daiteke, atributuak indibidualki erabilita, asmatze-tasa desberdinak lortu direla, baxuenetik altuenera: *bounding boxak* (0,2), hanketako angeluak (0,24), desplazamendu abiadura (0,31), belarri/esku distantziak (0,38) eta besoetako angeluak (0,48). Ondoren, ordena berean atributuak konbinatu dira, eta emaitza hobetuz joan da, atributu guztiekin 0,62 lortu den arte.

Atributu guztiak erabilita, baina eskeletoei arestian azaldutako leunketa aplikatu gabe, 0,61-eko asmatze-tasa lortu da, eta, leunduekin alderatuz, emaitza pixka bat soilik jaisten dela ikus daiteke, alegia, datu-leunketak eragin positibo oso txikia izan duela.

Bestalde, hurbilen dauden pertsonen %75 soilik hartzean, asmatze-tasa 0,61era jaitsi da, baina, %50 hartuz gero, 0,68ra igotzen da. Beraz, ereduak hobeto funtzionatzen du hurbilen dauden pertsonekin, urrutien daudenekin baino, pertsonen erdiak baztertzan direnean. Esan bezala, hori berresteko, pertsona guztiekin eta atributu guztiekin egin den pro-

ban, zeinetan 0,62ko asmatze-tasa lortu den, asmatutakoen eta ez-asmatutakoen *bounding boxen* altuerak kalkulatu dira. Asmatutakoen batez besteko altuera 274,25 pixelekoa izan da, eta ez-asmatutakoena, aldiz, 249,36 pixelekoa. Honekin azpimarratu daiteke asmatutakoetan batez besteko altuera nabarmen handiagoa dela eta, ondorioz, pertsonak hurbilago daudela kameratik.

Azkenik, frogatu nahi da ea gutxienez eskeletoa bere horretan erabiltzea bezain eraginkorra den lortutako atributuak erabiltzea; izan ere, arestian aipatu bezala, atributuak erabiltzean sortzen den neurona-sarea sinpleagoa da, eskeletoa bere horretan erabiltzen denean eraikitzen dena baino. Eskeletoa bere horretan erabiltzean, 0,64ko asmatze-tasa lortu da, hau da, atributuekin lortu den 0,62koa baino handiagoa. Horregatik, esan daiteke ez dela guztiz lortu neurona-sare sinpleago batekin emaitzak hobetzea, baina bai ia berdintzea.

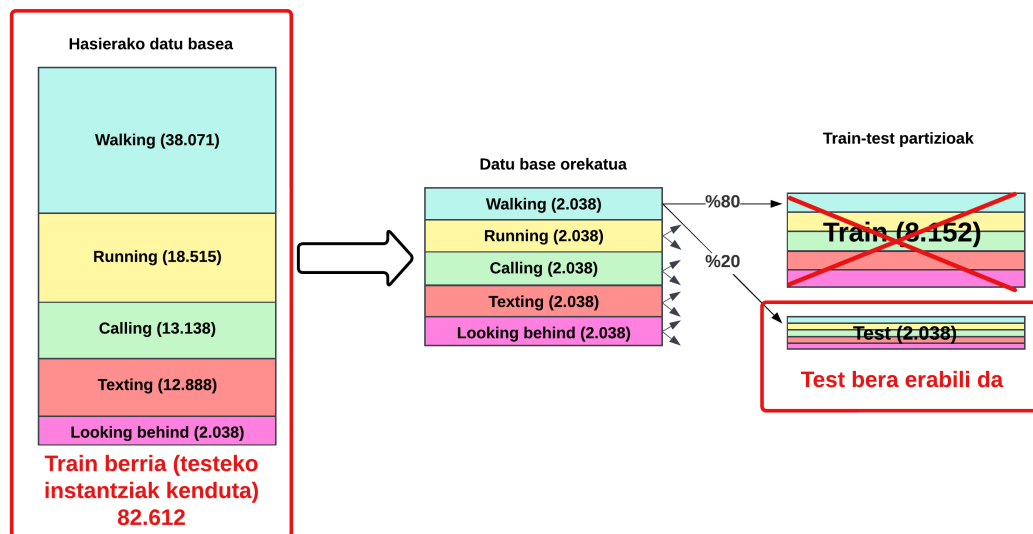
Datuak konparatu ostean, exekuzioetako bestelako datu batzuk ikusiko dira. Horretarako, datu leunduekin, atributu guztiak eta pertsonen %100 kontuan hartu den exekuzioa erabiltzeko da (0,62ko asmatze-tasa duena). Lehenengo, exekuzio horretako konfusio-matrizea aztertuko da, 4.11 Irudian ikus daitekeena.

| Konfusio matrizea | | | | | | |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---------------------------|
| TARGET \ OUTPUT | Atzera begira | Deika | Korrika | Txateatzen | Oinez | SUM |
| Atzera begira | 91 5% | 10 1% | 45 2% | 11 1% | 223 12% | 380 24% 76% |
| Deika | 0 0% | 315 17% | 4 0% | 18 1% | 43 2% | 380 83% 17% |
| Korrika | 1 0% | 7 0% | 286 15% | 50 3% | 36 2% | 380 75% 25% |
| Txateatzen | 4 0% | 21 1% | 86 5% | 198 10% | 71 4% | 380 52% 48% |
| Oinez | 47 2% | 27 1% | 18 1% | 3 0% | 285 15% | 380 75% 25% |
| SUM | 143 64% 36% | 380 83% 17% | 439 65% 35% | 280 71% 29% | 658 43% 57% | 1175 / 1900 62% 38% |

4.11 Irudia: Konfusio-matrizea.

Ekintzaz ekintza aztertuz gero, “Pertsona atzera begira”-rekin hasita, ikus daiteke asmatze-tasa baxua dela; batik bat, ekintza honetarako espreski atributurik ez delako erauzi izan daiteke. “Pertsona deika”-rekin jarraituz, emaitza hobetzen da atzera begira daudenekiko, beharbada ekintza honetarako atributu bat erauzi delako espreski. “Pertsona korrika”-k “Pertsona deika”-ren antzeko ildo jarraitzen du, eta arrazoi bertsuegatik izan daiteke. “Pertsona txateatzen” ekintzan, emaitzak jaitsi egiten dira pixka bat, “Pertsona korrika” eta “Pertsona oinez” ekintza gisa sailkatzen baititu okerrak, eta, neurri txikiagoan, “Pertsona deika” gisa. Azkenik, “Pertsona oinez” ekintzan asmatze-tasa handia da, baina en-gainatzailea, izan ere, ekintza guztietan “Pertsona oinez” bezala sailkatzeko joera handia du ereduak. Hain zuzen ere, oinezkotzat hartzen dituen 658 pertsonetatik %43 bakarrik asmatzen du. Beraz, esan daiteke instantziak “Pertsona oinez” gisa sailkatzeko ohitura edo desbideraketa duela ereduak.

Beste konparaketa bat gehiago egin da. Zehazki, *tester*ako instantzia berak mantendu dira, baina entrenamendurako hasiera-hasieran orekatu aurretik zegoen datu-base osoa erabili da (*tester*ako erabiliko diren instantziak kenduta, noski) (4.12 Irudia).



4.12 Irudia: Train eta test partizioen banaketa berria.

Hori eginda, asmatze-tasa 0,75era igo da, eta 4.13 Irudian ikus daitekeen konfusio-matrizea lortu da.

| Konfusio matrizea | | | | | | |
|-------------------|------------------|------------------|-------------------|-------------------|-------------------|---------------------------|
| TARGET \ OUTPUT | Atzera begira | Deika | Korrika | Txateatzen | Oinez | SUM |
| Atzera begira | 34 2% | 13 1% | 45 2% | 62 3% | 226 12% | 380 9% 91% |
| Deika | 1 0% | 344 18% | 3 0% | 13 1% | 19 1% | 380 91% 9% |
| Korrika | 0 0% | 1 0% | 339 18% | 18 1% | 22 1% | 380 89% 11% |
| Txateatzen | 5 0% | 3 0% | 14 1% | 342 18% | 16 1% | 380 90% 10% |
| Oinez | 6 0% | 3 0% | 7 0% | 0 0% | 364 19% | 380 96% 4% |
| SUM | 46 74% 26% | 364 95% 5% | 408 83% 17% | 435 79% 21% | 647 56% 44% | 1423 / 1900 75% 25% |

4.13 Irudia: Konfusio-matrizea *train* handituz.

Begi-bistakoa da emaitzak hobeak direla, batez ere, “Pertsona oinez” bezala sailkatzearen desbideraketa txikitu delako, ziurrenik, oinezkoen askoz datu gehiago izanda, errazago desberdintzen dituelako oinezkoak eta bestelako ekintzak. Entrenamendu-multzoan “Atzera begira” dauden instantziak ez dira areagotu, instantzia gutxien dituen ekintza baitzen lehen ere; horregatik, emaitzak ez dira hobetu. Aldiz, gainontzeko lau ekintzetan asmatze-tasak askoz hobeak dira. Ondorioz, esan dezakegu *train* multzoa handituz hobekuntza nabarmena izan dela eta, instantzia berberak izanda ere, *testean* daudenak, hobeto sailkatzen dituela.

Datu hauek guztiak aztertu ostean, esan daiteke asmatze-tasak ez direla estandar onargarrietara iristen. Hala ere, atributuen erabilerari dagokionez, horrelako ereduak eraikitzean, aintzat hartu daitezkeen konklusio baliagarriak ondorioztatu daitezke lortutako datuetatik. Ondorioen atalean (5) jasoko dira konklusio hauek guztiak.

5. KAPITULUA

Ondorioak

Lan honen garapenean zehar eta, batez ere, amaitu ostean, ondorio desberdinak jaso dira. Horietako batzuk lan zehatz honekin dute zerikusia; besteek, ekintza-identifikazioarekin, aplikazioekin, etikarekin... Guztiak jasoko dira atal honetan.

Hasteko, lan zehatz honi dagokionez, datu desberdinen konparaketatik emaitza eta ondorio desberdinak lortzearen balioa ikusi da. Esaterako, kameretatik urrutien dauden pertsonen ekintzen identifikazioa gertukoena baino txarragoa dela ikusi daiteke datuetan. Bestela esanda, asmatu ez diren pertsonen batez besteko altuera txikiagoa da (urruti dauden seinale), asmatutakoena baino, eta, gertuen daudenak soilik hautatuz joan garen neurrian, emaitzak hobetu dira. Hortik ondoriozta daiteke lehenengo kamerak urrutiko pertsonak arbuiatu ditzakeela eta bigarren kamera bat ezarri litekeela urruti dauden horietatik gertuago.

Horrez gain, lan honen helburuetako bat izan da, datu gutxiago baina esanguratsuagoak erabilita, emaitza antzekoak lor daitezkeen ala ez ondorioztatzea, hau da, eskeleto osoak erabili beharrean, bestelako atributuak baliatzea. Emaitzen interpretazioan ikusi bezala, eredu honetan ia lortu da emaitzak berdintzea atributu gutxi baina proportzionalki esanguratsuagoekin. Konparaketa positibo hau ikusita eta oraindik askoz atributu gehiago erauzi daitezkeela kontuan hartuta, jorratu beharreko alderdi bat dela ondorioztatu da. Bestalde, eskeletoen leunketak eragin positiboa duela ondorioztatu ahal izan da, baina honekin eta abiadura eta antzeko atributuekin kontuan izan behar da fotograma bat baino gehiagotako informazioa eskatzen dutela eta denbora errealeko aplikazioak egiteko ez direla beti erabilgarri. Hain zuzen ere, aurreko fotograma gutxi batzuk soilik erabiltzea aproposa izan daiteke denbora errealeko aplikazioetan, baina ezin izango dira fotograma gehiegi erabili emaitza bat denbora errealean eman nahi bada.

Ekintza-identifikazioaren arloan, edo orokorrean ikusmen artifizialean, ondorio orokorrago batzuk ere atera daitezke. Hasteko, edozein eredu garatzen ibilita ere, datu asko izatea ezinbestekoa da ereduak entrenatzeko. Ildo berean, aipatu beharra dago, datu asko tratatzeko eta potentzia handiko eredu bat trebatzeko, ordenagailu simple bat baino gehiago behar dela. Hain zuzen ere, lan hau adibide gisa hartuta, emaitza kaskarrak lortu arren, ikerketa ez da nahikoa eskeletoen erabilera ez dela esanguratsua esateko eta lortu diren

atributuen erabilera baztertzeko. Are gehiago, hiru ezkutuko geruzako eredu soil bat entrenatu da lan honetan, eta konparazioak egiteko baliagarria izan da, baina ez da arlo honetan egin daitezkeen hipotesiak baztertzeko adinakoa.

Gainera, ikusmen artifizialaren arloa, berria ez ezik, zabala dela ikus daiteke, teknika berriak etengabe sortzen aritzen direlako eta baita ikusmen artifiziala aplikatu daitezkeen testuinguruak asko direlako ere: jende-multzoak, pertsonak banaka, ekintza desberdin anitz, ondorioztatu daitezkeen aplikazio desberdinak. . . Bestela esanda, eta berriro ere lan honetan ikusi dena itzulita, aparkaleku batean gertatu dena aztertu ordez, beste kokaleku bat aztertu zitezkeen, edota bestelako ekintzak, edota jende gehiagoko multzoak, edota jende-multzoen portaerak. . . Hitz gutxitan, ezaugarri desberdinetako hamaika lan burutu daitezke ikusmen artifizialaren barruan.

Etorkizunera begira, aparkalekuetako segurtasuna hobetzeko, ereduak hobetu beharko dira, eta, horrela, autoetako, aparkalekuetako edo kaleetako segurtasun kamerak erabili ahal izango dira, ekintzak detektatzeko eta erantzun gisa auto adimendunek erabakiak hartzeko, arriskuak saihestuz. Erabaki horiek hartzeko, alde etiko bat ere jorratu behar dela argi dago, erabakiak hartzea makinaren esku uzten den momentutik.

Ereduak hobetzeari dagokionez, lan honetan landu diren atal desberdinetatik abiatuz, jarraipena emateko aukera asko ondorioztatu daitezke. Hasteko, arestian aipatu bezala, geruza gehiagoko neurona-sareak ez ezik, bestelako teknikak erabili daitezke edota teknika desberdinen konbinaketak. Esaterako, eskeletoetatik atributu gehiago erauzteaz gain, sare neuronal konboluzionalak erabili daitezke [Tasnim and Baek, 2023] edota biekin elkarlanean aritu. Bestalde, datu-base gehiago erabilita, emaitzak hobetzen diren aztertu behar litzateke, baita aurreprozesaketa fasean proba berriak egin ere (datuen leunketa desberdinak, kasu). Hitz gutxitan, ekintzen sailkapenerako emandako pauso bakoitzean, badago zer egin oraindik.

Azkenik, ikusmen artifiziala etorkizun hurbilean *deep learning*aren bidetik jorratuko dela ondorioztatu daiteke [Basheer Ahmed et al., 2023], neurona-sare konboluzionalen garapenarekin, adibidez. Objektuen detekzioan, irudien segmentazioan, sailkapenean. . . gertatuko dira garapenak, eta mundu errealeko arazo eta arloetan aplikatu ahal izango dira emaitzak, erronka berriak sortuz eta gaituz.

Bibliografia

- [can, 2023] (2023). Canva. <https://www.canva.com>. Accessed June 11, 2023.
- [Basheer Ahmed et al., 2023] Basheer Ahmed, M. I., Zaghdoud, R., Ahmed, M. S., Sendi, R., Alsharif, S., Alabdulkarim, J., Albin Saad, B. A., Alsabt, R., Rahman, A., and Krishnasamy, G. (2023). A real-time computer vision based approach to detection and classification of traffic incidents. *Big data and cognitive computing*, 7(1):22.
- [Bouwman et al., 2018] Bouwman, T., Silva, C., Marghes, C., Zitouni, M. S., Bhaskar, H., and Frelicot, C. (2018). On the role and the importance of features for background modeling and foreground detection. *Computer Science Review*, 28:26–91.
- [Cao et al., 2018] Cao, Z., Hidalgo, G., Simon, T., Wei, S., and Sheikh, Y. (2018). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008.
- [Cao et al., 2019] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., and Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [Cao et al., 2017] Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*.
- [Connaught Automotive Research (CAR) Group, 2005] Connaught Automotive Research (CAR) Group (2005). <https://car.universityofgalway.ie/>.
- [DimitrisMi, 2023] DimitrisMi, T. u. (2023). Neural network convolutional neural networks gif. Online. Accessed on June 4, 2023.
- [Freeman and Roth, 1995] Freeman, W. T. and Roth, M. (1995). Orientation histograms for hand gesture recognition. In *International workshop on automatic face and gesture recognition*, volume 12, pages 296–301. Citeseer.
- [Harris et al., 2020] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585:357–362.

- [Haykin, 1994] Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.
- [Huang and Wu, 2020] Huang, Y.-F. and Wu, H.-Y. (2020). Image retrieval based on asift features in a hadoop clustered system. *IET Image Processing*, 14(1):138–146.
- [Jin et al., 2017] Jin, C., Li, S., and Kim, H. (2017). Real-time action detection in video surveillance using sub-action descriptor with multi-cnn. *CoRR*, abs/1710.03383.
- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [Liu and Nocedal, 1989] Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.
- [Liu et al., 2022] Liu, R., Ramli, A. A., Zhang, H., Henricson, E., and Liu, X. (2022). An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence. In *Internet of Things–ICIOT 2021: 6th International Conference, Held as Part of the Services Conference Federation, SCF 2021, Virtual Event, December 10–14, 2021, Proceedings*, pages 1–14. Springer.
- [Lowe, 1999] Lowe, D. (1999). Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157 vol.2.
- [McConnell, 1986] McConnell, R. K. (1986). Method of and apparatus for pattern recognition.
- [McKinney et al., 2010] McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.
- [Ojala et al., 1994] Ojala, T., Pietikainen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 582–585 vol.1.
- [Park et al., 2005] Park, H. S., Kim, E. Y., Jang, S. S., and Kim, H. J. (2005). Recognition of human action for game system. In *Artificial Intelligence and Simulation: 13th International Conference on AI, Simulation, Planning in High Autonomy Systems, AIS*

2004, Jeju Island, Korea, October 4-6, 2004, Revised Selected Papers 13, pages 100–108. Springer.

- [Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- [Rautaray and Agrawal, 2015] Rautaray, S. S. and Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial intelligence review*, 43:1–54.
- [Raybaut, 2009] Raybaut, P. (2009). Spyder-documentation. Available online at: *pythonhosted.org*.
- [Rosebrock, 2016] Rosebrock, A. (2016). Intersection over union (iou) for object detection.
- [Ruder, 2016] Ruder, S. (2016). An overview of gradient descent optimization algorithms.
- [Saturn Cloud, 2018] Saturn Cloud (2018). A Comprehensive Guide to Convolutional Neural Networks: The ELI5 Way. <https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/>. Accessed: 06 11, 2023.
- [Simon et al., 2017] Simon, T., Joo, H., Matthews, I., and Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*.
- [Simonyan and Zisserman, 2015] Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.
- [Tasnim and Baek, 2023] Tasnim, N. and Baek, J.-H. (2023). Dynamic edge convolutional neural network for skeleton-based human action recognition. *Sensors*, 23(2):778.
- [Van Rossum and Drake Jr, 1995] Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
- [Wei et al., 2016] Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y. (2016). Convolutional pose machines. In *CVPR*.

[Zheng et al., 2018] Zheng, Y., Yang, C., and Merkulov, A. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. page 4.