Fakulta informačních technologií
Vysoké učení technické v Brně

Strojové učení a rozpoznávání
Identifikace osob z obrázku obličeje a hlasové nahrávky

5. května 2025

Anna Ovesná

xovesn03

1 Úvod

Cílem tohoto projektu bylo vyvinout systém pro identifikaci osob na základě analýzy obrázků obličeje a zvukových nahrávek. Tento systém využívá pokročilé metody strojového učení, zejména konvoluční neuronové sítě (CNN) pro zpracování obrazových dat a techniky zpracování signálu. U obou typů dat bylo zapotřebí rozšířit sadu augmentací pro efektivnější trénování.

2 Identifikace z obrázku

Pro rozpoznávání obličejů na obrázcích byla použita konvoluční síť, běžně využívaná pro analýzu obrazových dat. Model se skládá z konvolučních vrstev, které extrahují rysy z obrázku, a pooling vrstev, které redukují rozměry a zajišťují nižší výpočetní nároky. Proces je doplněn vrstvami normalizace stabilizujícími trénování a aktivační funkcí ReLU, která modeluje nelineární závislosti v datech. K prevenci přetrénování byl použit dropout, který náhodně vypíná určité neurony během trénování.

Model obsahuje čtyři konvoluční vrstvy postupně s 64-64-128-32 výstupními neurony, z nichž pouze po první následuje pooling. Na závěr je přidána plně propojená vrstva s 55 neurony pro klasifikaci na základě extrahovaných rysů. Tato architektura vznikla iterativním experimentováním.

Před zpracováním dat probíhá normalizace a augmentace obrázků. Data jsou normalizována pomocí skutečného průměru a směrodatné odchylky vždy z trénovacích dat, což zajišťuje konzistentní měřítko vstupních dat a lepší výkon modelu.

Pro augmentaci obrázků byly využity transformace, které zajišťují lepší generalizaci modelu a snižují riziko přetrénování. Každá z těchto metod byla vybrána experimentálně. Použité metody jsou implementovány v knihovně PyTorch a zahrnují:

- **RandomHorizontalFlip**: Náhodně otáčí obrázek kolem horizontální osy. Pomáhá modelu lépe generalizovat v případě, kdy je objekt na obrázku umístěn v různých orientacích.
- **RandomRotation**: Rotuje obrázek o určité úhly. Tento krok modelu umožňuje naučit se identifikovat objekty bez ohledu na jejich konkrétní orientaci.
- **RandomResizedCrop**: Náhodně mění velikost obrázku.
- **ColorJitter**: Mění kontrast, sytost a jas obrázku. Pomáhá modelu rozpoznávat objekty i za různých světelných podmínek a v proměnlivých kvalitách snímků.

Testovací sada použita během vývoje dosáhla průměrné přesnosti 83.25 % a maximální přesnosti 93.55 % na 20 nezávislých bězích.

3 Identifikace ze zvukové nahrávky

Preprocessing zvukových nahrávek zahrnuje aplikaci několika augmentačních technik, které mají za cíl rozšířit trénovací vzorky a zlepšit schopnost modelu generalizovat. Po načtení trénovací nahrávky byly s určitou pravděpodobností provedeny následující augmentace:

- **TimeMasking**: Náhodné skrytí část signálu v časovém prostoru, což pomáhá modelu zaměřit se na jiné části signálu.
- **FrequencyMasking**: Skrytí určitých frekvenčních komponent zvukového signálu, což umožňuje modelu naučit se ignorovat šum nebo neúplné signály.
- **TimeStretch**: Náhodná změna délky zvukového signálu v čase, což simuluje různé rychlosti mluvy nebo zpěvu.
- **Spektrogramová transformace**: Změna struktury signálu ve frekvenční doméně a následně převod zpět do časové domény.

Po případné augmentaci byla aplikována technika Mel-frequency cepstral coefficients (MFCC), která extrahuje akustické rysy důležité pro rozpoznávání zvuku. Z MFCC byly extrahovány statistické vlastnosti (průměr, směrodatná odchylka, maximum a minimum), čímž došlo k redukci dimenze a usnadnění použití jednodušších modelů. Tímto způsobem byly vytvořeny nové trénovací vzorky, které prošly následnou normalizací.

Zvukové nahrávky byly opakovaně zpracovány různými způsoby, což vedlo k rozšíření trénovacích dat a lepší adaptabilitě modelu na reálné podmínky.

Pro klasifikaci zvukových nahrávek byla zkoušena řada metod strojového učení počínaje Gaussian Mixture Model (GMM). Ačkoli se tato metoda doporučuje pro podobné úkoly, její výkon se ukázal být velmi nízký, dosahující přesnosti pouze kolem 3 %. Tento výsledek naznačuje, že GMM není vhodná pro tento konkrétní způsob předzpracování dat.

Mezi další testované metody patří K-nejbližších sousedů (2 - 10 sousedů) a Random Forest (desítky až stovky estimators). Obě tyto metody vykazovaly lepší, ale stále nedostatečnou přesnost, která se pohybovala ne nikdy na více než kolem 40 %. Tyto metody poskytly určité zlepšení oproti GMM, ale stále ne dostatečně pro úkol identifikace osob.

Nejlepších výsledků dosahoval Support Vector Machine (SVM) s RBF (Radial Basis Function) jádrem. Přesnost vzrostla na podstatně vyšší hodnoty, než jaké byly dosaženy s předchozími metodami a to tak, že během vývoje bylo na testovací sadě dosaženo průměrné přesnosti 78.47 % a maximální přesnosti 82.26 %.

4 Spuštění

Kód je rozdělen do několika souborů a tříd, které se zaměřují na konkrétní úkoly, jako je preprocessing dat a samotné modelování.

Složka `src` obsahuje veškeré spouštěcí skripty a k jejich správnému běhu je zapotřebí mít nainstalované balíčky ze souboru `requirements.txt`.

Pro spuštění kódu je třeba nejprve aktivovat virtuální prostředí a ujistit se, že jsou všechny závislosti nainstalovány. Následně lze spustit jeden z hlavních souborů. Soubor `identify_png.py` je určen pro práci s obrázky ve formátu PNG a `identify_wav.py` pro zvukové soubory ve formátu WAV. Každý soubor spustí opakovaně trénování modelu, aby našel ten s nejlepší přesností, který následně uloží.

Pro jednoduchost načítání evaluačních dat byla manuálně do složky `eval` vnořena podsložka simulující třídu všech dat. Ve skriptu `evaluate.py` na řádku 19 a 20 lze změnit požadovaný typ zpracovávaných dat. Následně dojde k přípravě dat pro evaluaci, samotné vyhodnocení a uložení výsledků do souboru.