A new base-derivative framework is proposed, where base refers to the original visible and infrared modalities, and derivative refers to the two auxiliary modalities that are derived from base.

In the proposed framework, the double-modality cross-modal learning problem is reformulated as a four-modality one. After that, the images of all the base and derivative modalities are fed into the feature learning network. With the doubled input images, the learned person features become more discriminative.

Furthermore, the proposed framework is optimized by the enhanced intra- and cross-modality constraints with the assistance of two derivative modalities.

Three loss functions are used to guide the learning.

- ID loss
- Multi-mode intra-modality loss (MML)
  MML enlarges the general two-mode intra-modality triplet loss into multimode (two base modalities and two derivative modalities).
- Multi-directional cross-modality loss (MDL)

The four modalities take different streams (shown in different colors). The feature learner is divided into two parts: four modality-independent feature learners and a modality-sharing deep feature learner.

The four modalities take different streams (shown in different colors). The feature learner is divided into two parts: four modality-independent feature learners and a modality-sharing deep feature learner.