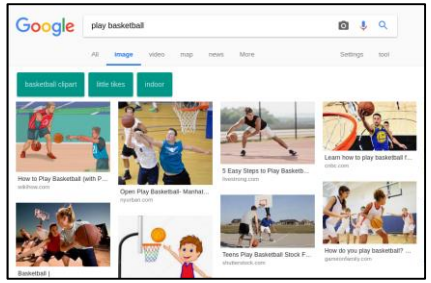
Learning from Heterogeneous Web Sources

Li Niu

Heterogeneous Web Sources

Heterogeneous web sources can be used as training data.

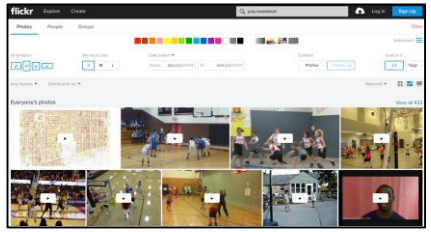
source domain 1:
Google images
and text



source domain 2:
Bing images
and text

•
•
•

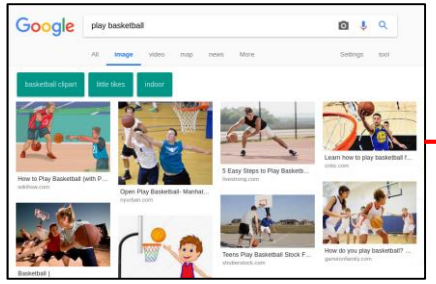
source domain *S*:
Flickr videos
and text



Domain Shift between Source Domains and Target Domain

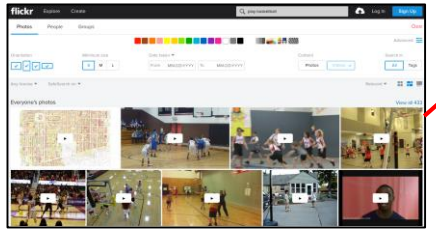
To address the domain shift, identify the most matched source domain for target domain.

source domain 1:
Google images
and text



source domain 2:
Bing images
and text

source domain *S*:
Flickr videos
and text

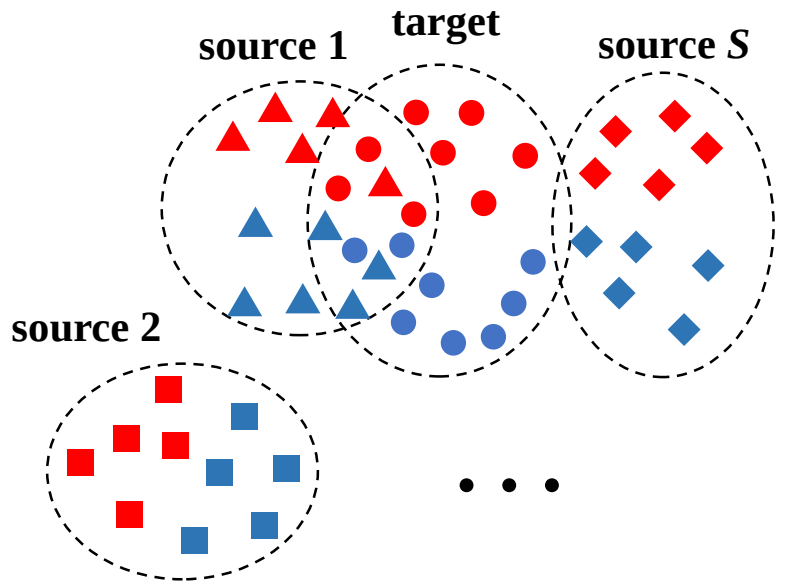


d^1

d^2

d^S

target domain



Flowchart of Our Method

Video Event Detection

source domain 1:
Google images
and text

Poland playing
Belgium in Qualifiers

Effects of Sleep on
Running Performance

source domain 2:
Bing images
and text

Youth Sports

Journal's new sports
photo archive

•
•
•

source domain *S*:
Flickr videos
and text

Final da Copa do
Brasil 2008

Live Spurs

2D visual features
textual features

2D visual features
textual features

3D visual features
textual features

Our
learning
model

Flowchart of Our Method

Video Event Detection

source domain 1:
Google images
and text


Poland playing Belgium in Qualifiers

Effects of Sleep on Running Performance

source domain 2:
Bing images
and text


Youth Sports

Journal's new sports photo archive

source domain *S*:
Flickr videos
and text


Final da Copa do Brasil 2008

Live Spurs

2D visual features
textual features

2D visual features
textual features

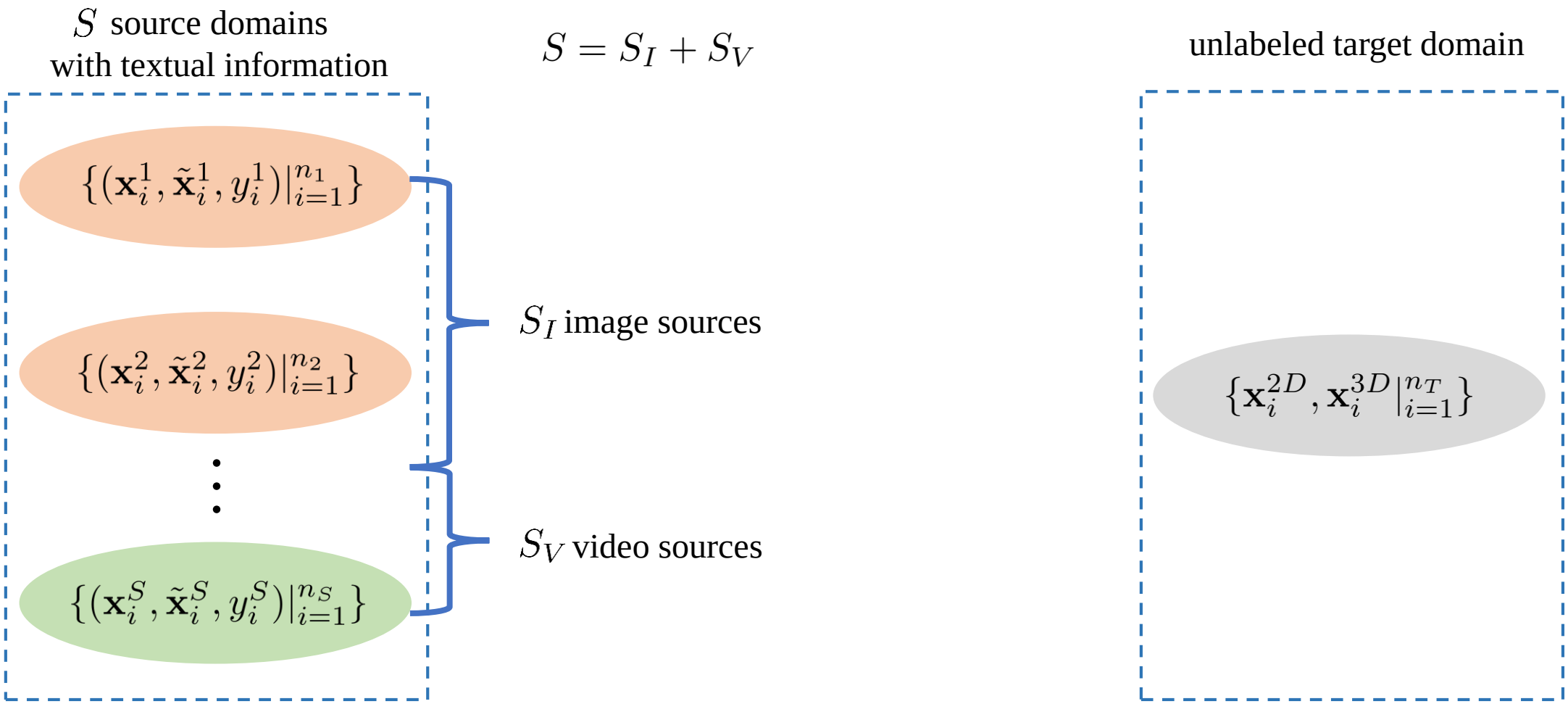
3D visual features
textual features

Our learning model

unlabeled target domain

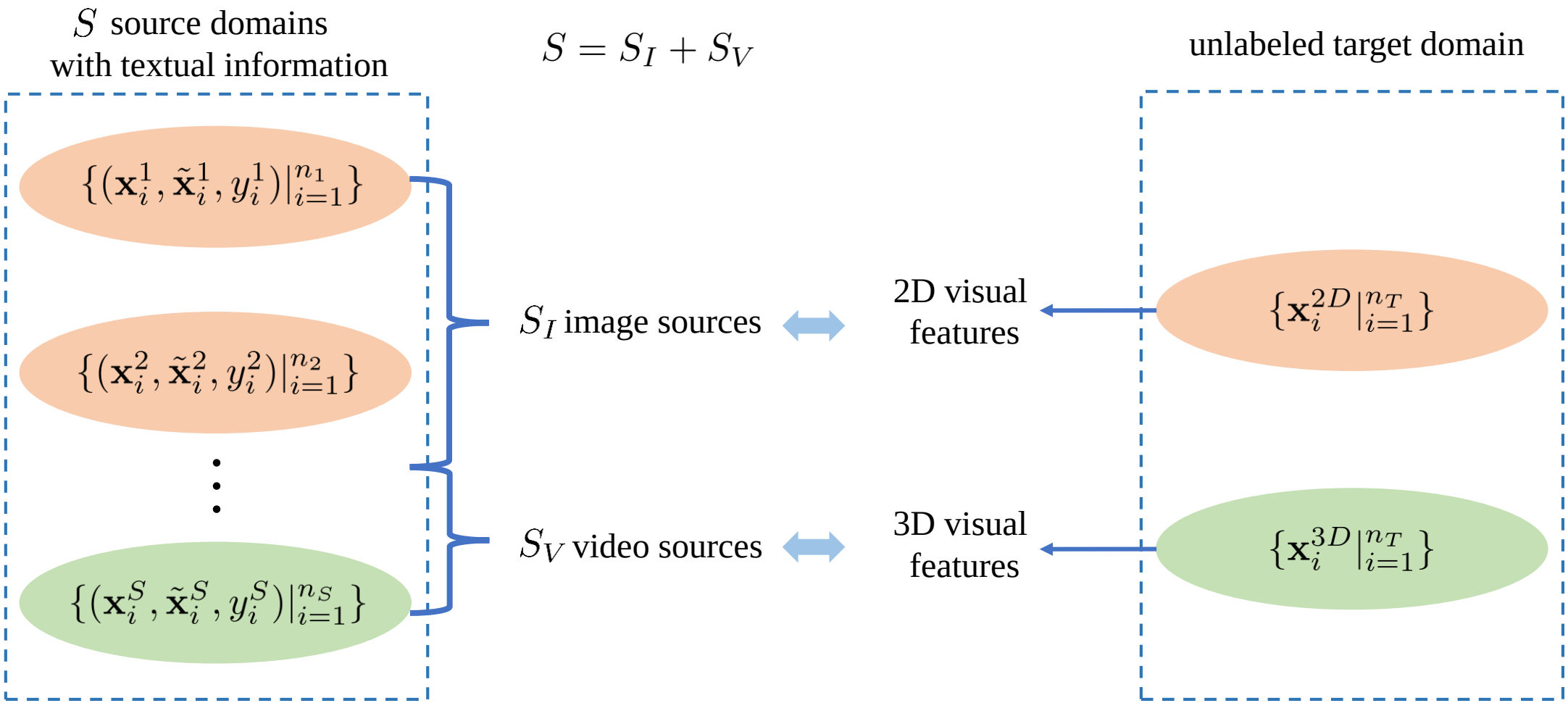

2D visual features
3D visual features

Formal Representation of Our Method



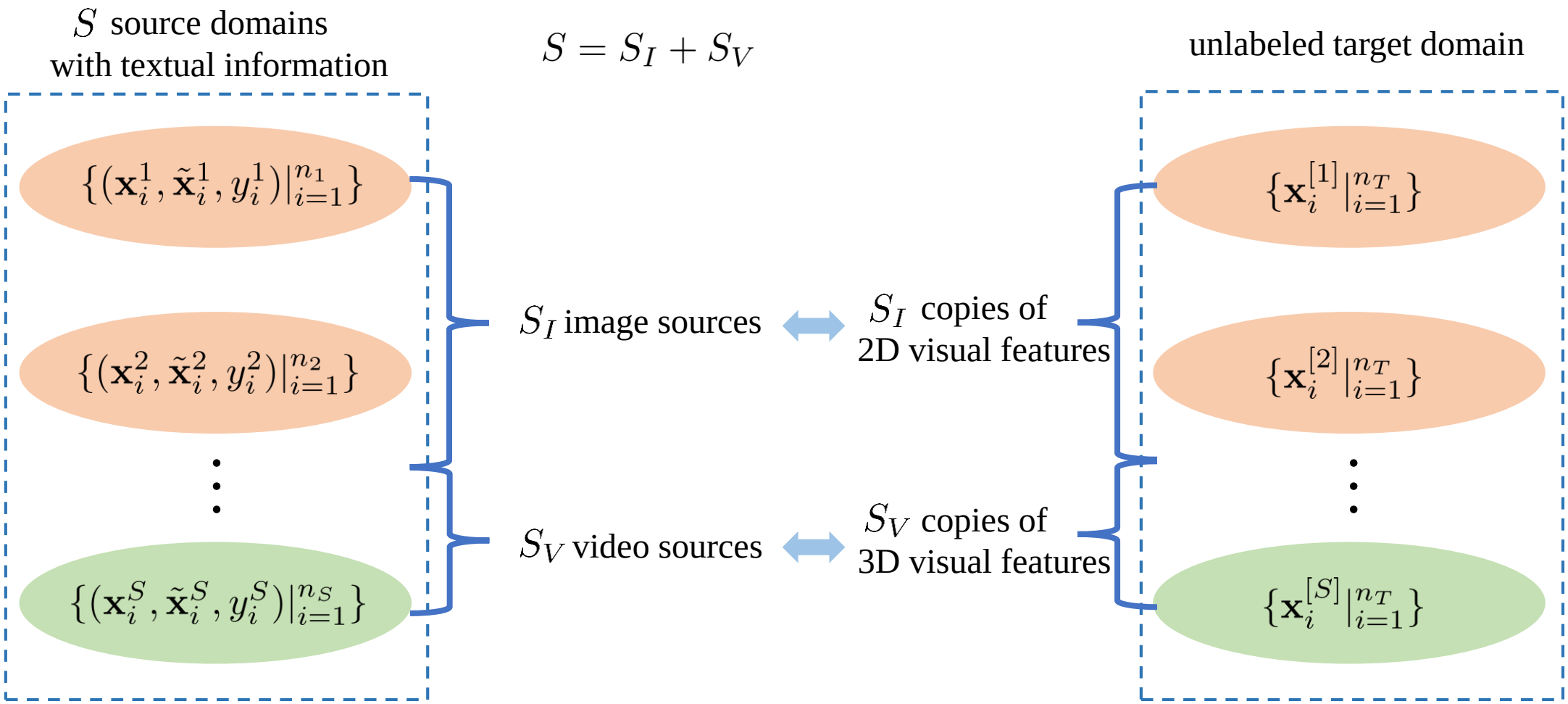
	visual	textual
feature	$\mathbf{x}_i^s$, $\mathbf{x}_i^{[s]}$	$\tilde{\mathbf{x}}_i^s$

Formal Representation of Our Method



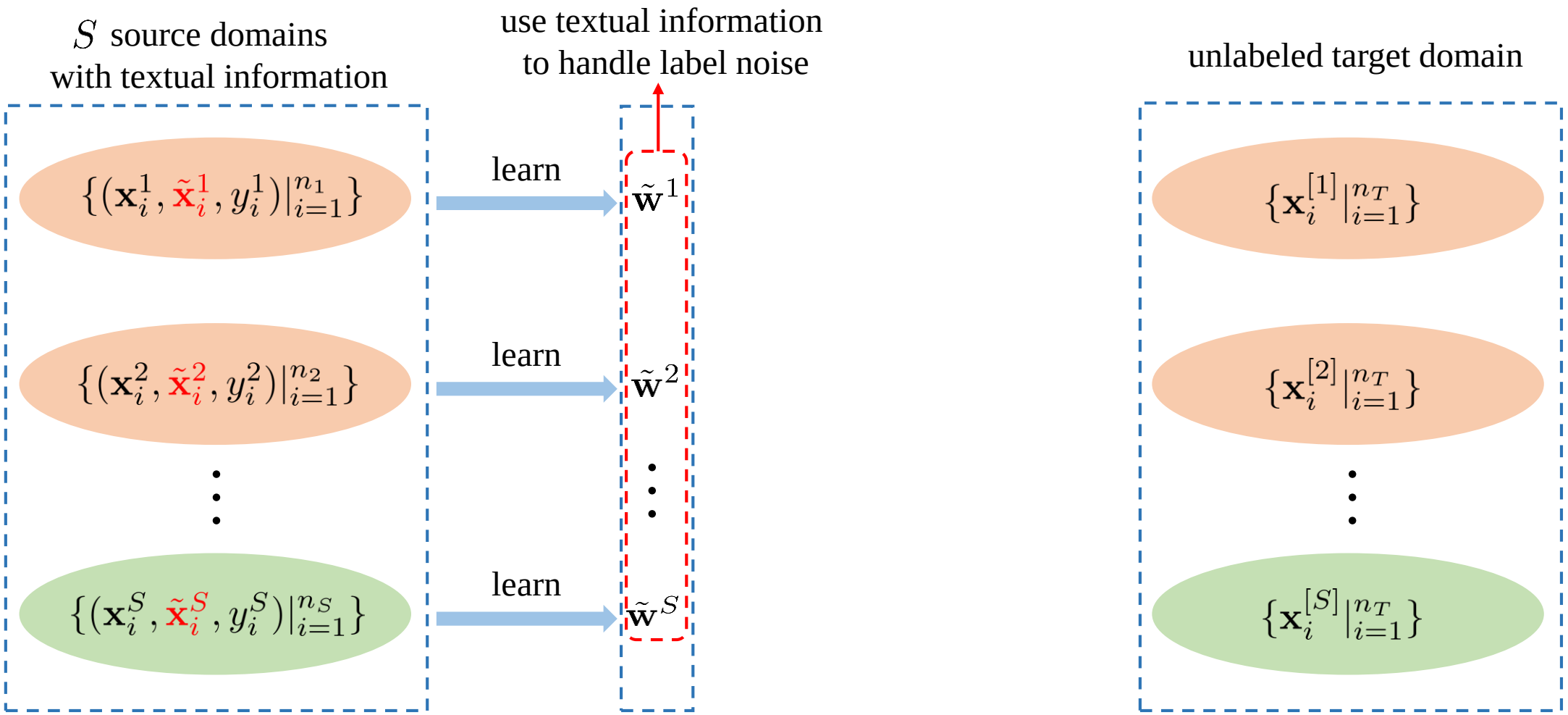
	visual	textual
feature	$\mathbf{x}_i^s$, $\mathbf{x}_i^{[s]}$	$\tilde{\mathbf{x}}_i^s$

Formal Representation of Our Method



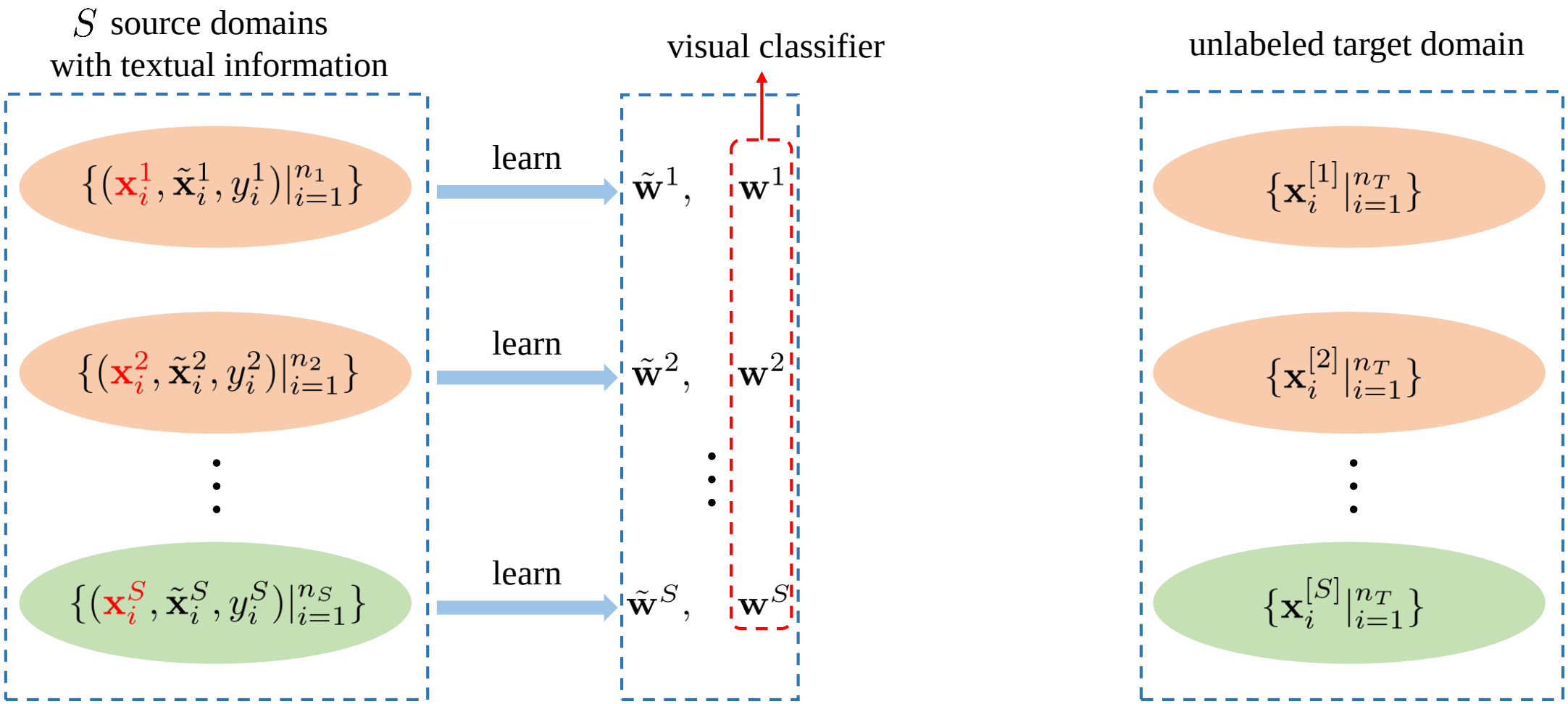
	visual	textual
feature	$\mathbf{x}_i^s$, $\mathbf{x}_i^{[s]}$	$\tilde{\mathbf{x}}_i^s$

Handle Label Noise



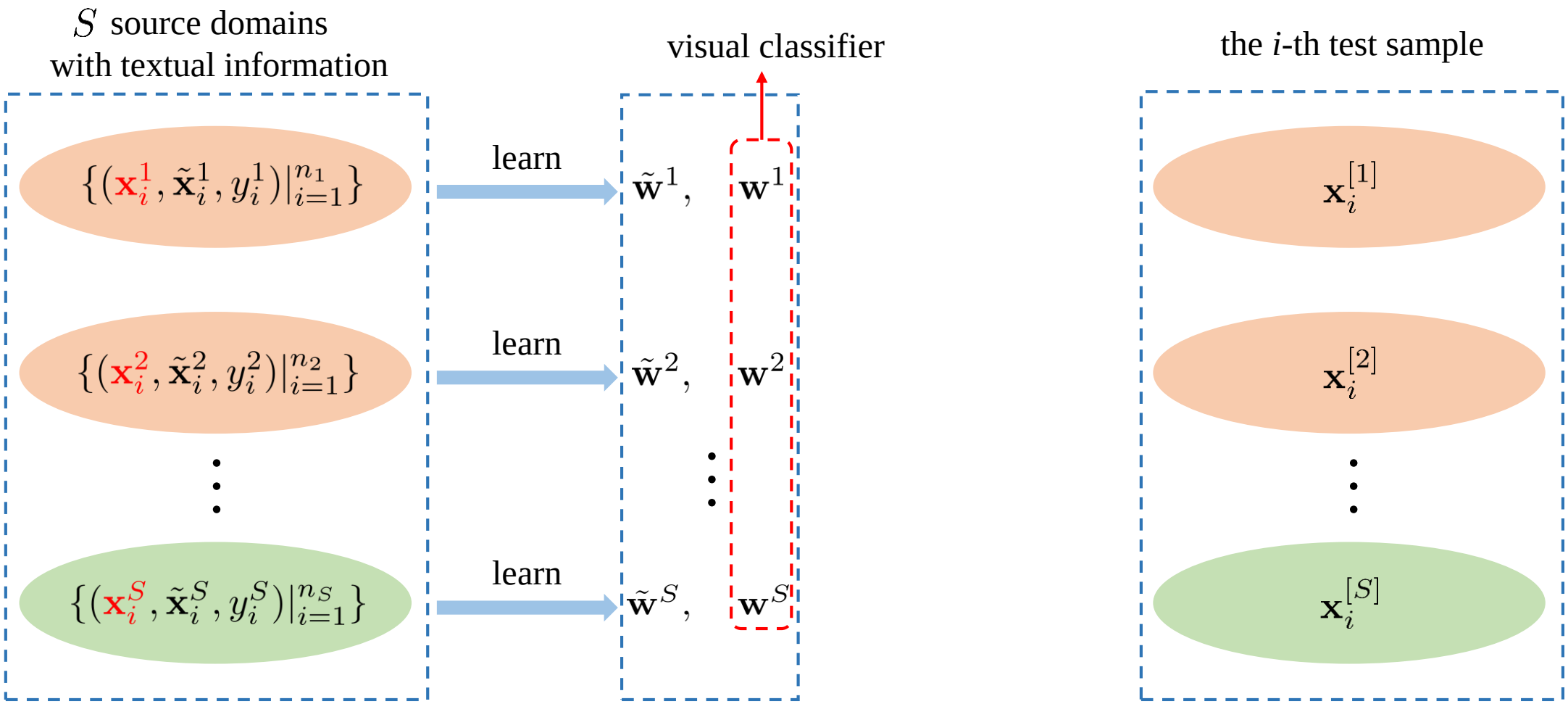
	textual	
model parameter	$\tilde{\mathbf{w}}^s$	

Learn Visual Classifier



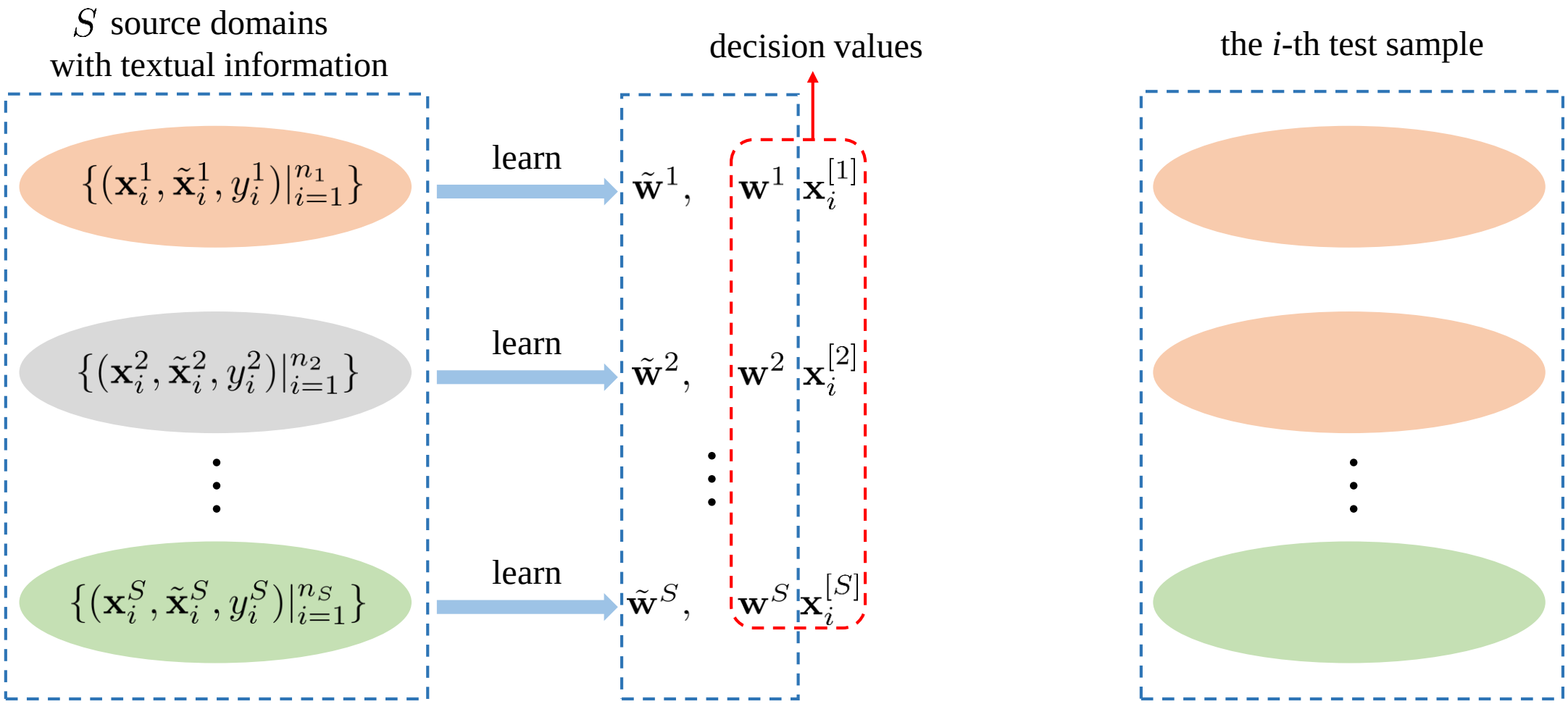
	textual	visual
model parameter	$\tilde{\mathbf{w}}^s$	$\mathbf{w}^s$

Learn Visual Classifier

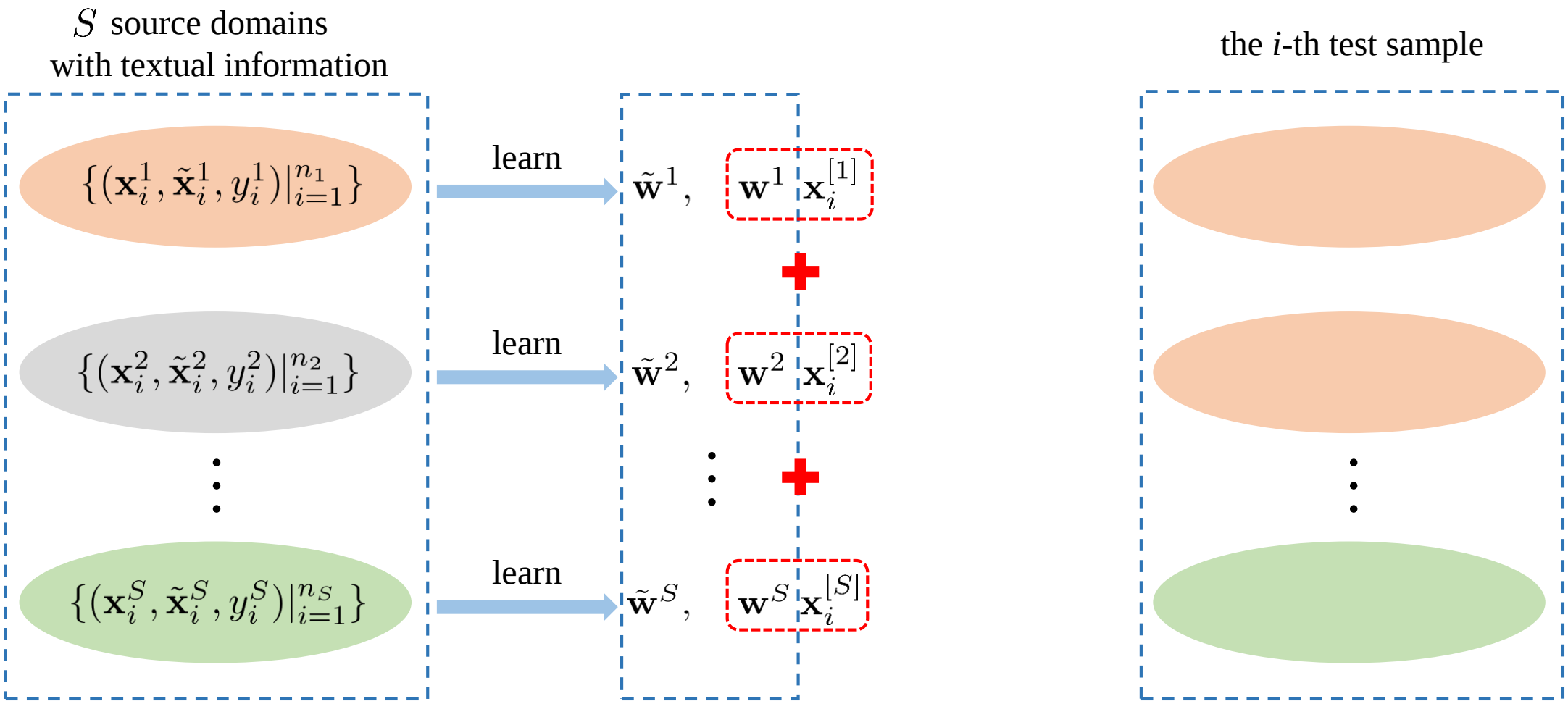


	textual	visual
model parameter	$\tilde{\mathbf{w}}^s$	$\mathbf{w}^s$

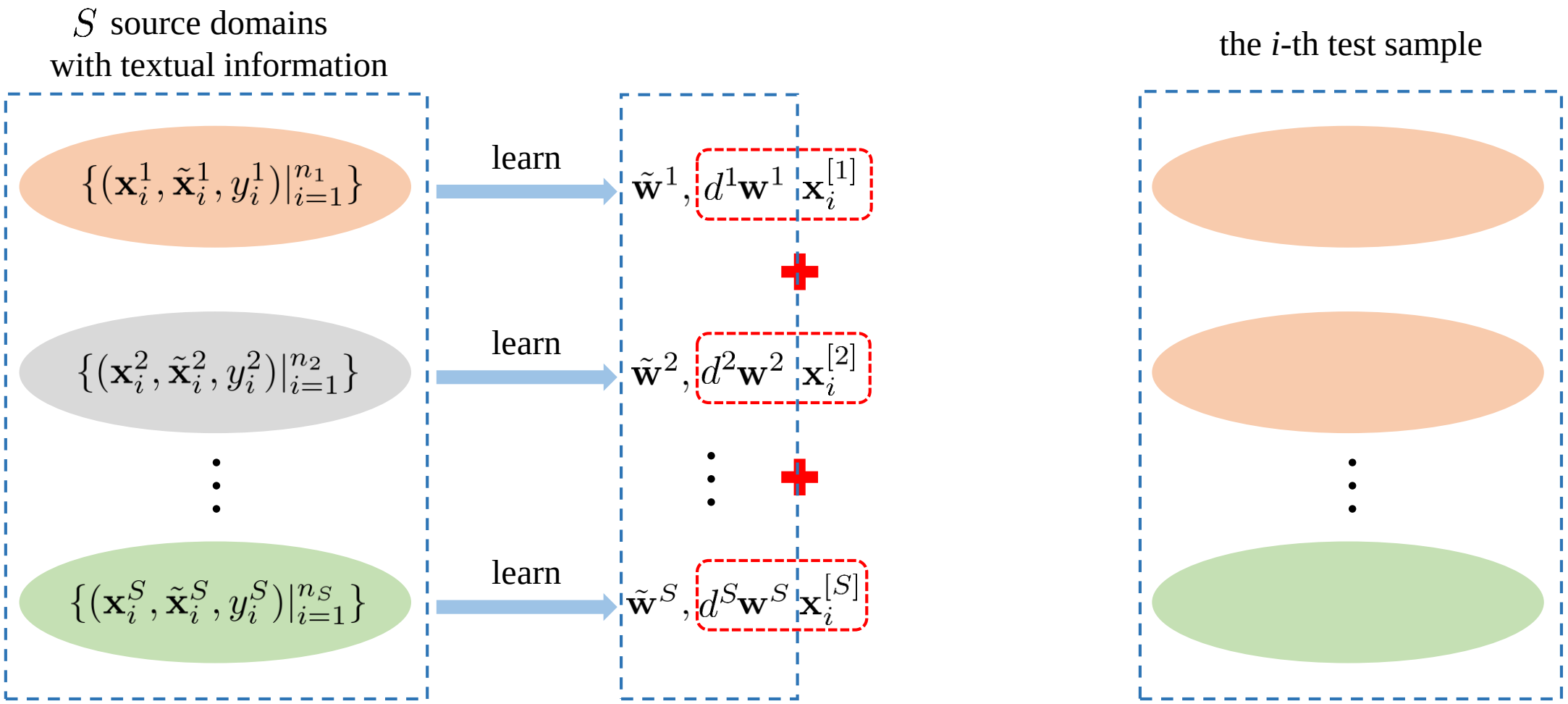
Learn Visual Classifier



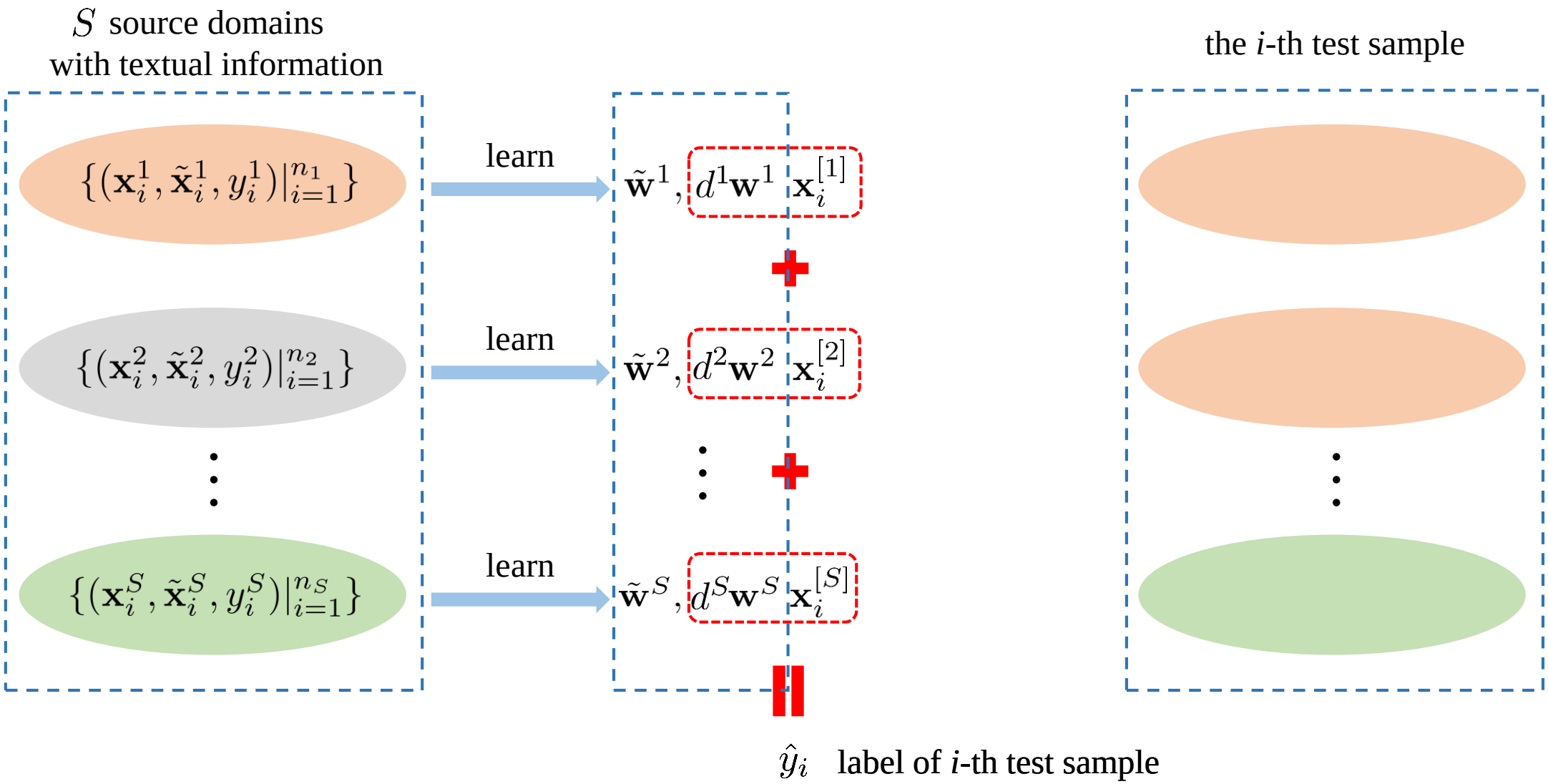
Learn Visual Classifier



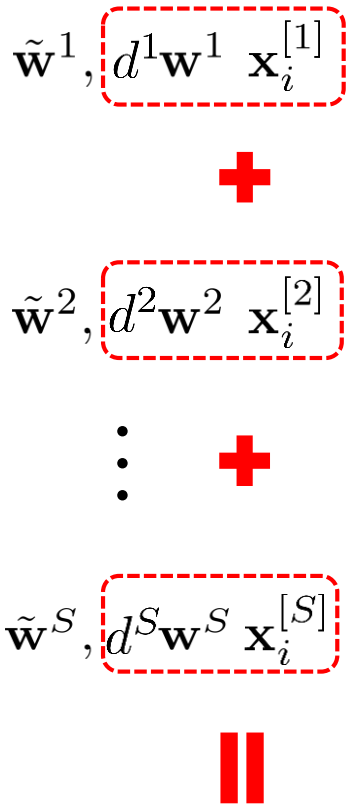
Learn Visual Classifier



Learn Visual Classifier



One Unified Approach



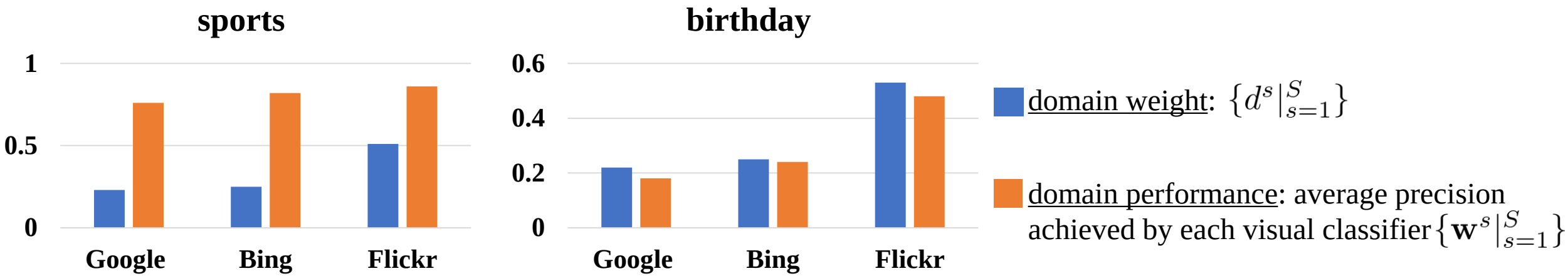
$\hat{y}_i$ label of *i*-th test sample

Our Approach

- learn $\{\mathbf{w}^s|_{s=1}^S\}$ and $\{\tilde{\mathbf{w}}^s|_{s=1}^S\}$
- learn domain weights $\{d^s|_{s=1}^S\}$
- infer the labels of test samples $\{\hat{y}_i|_{i=1}^{n_T}\}$

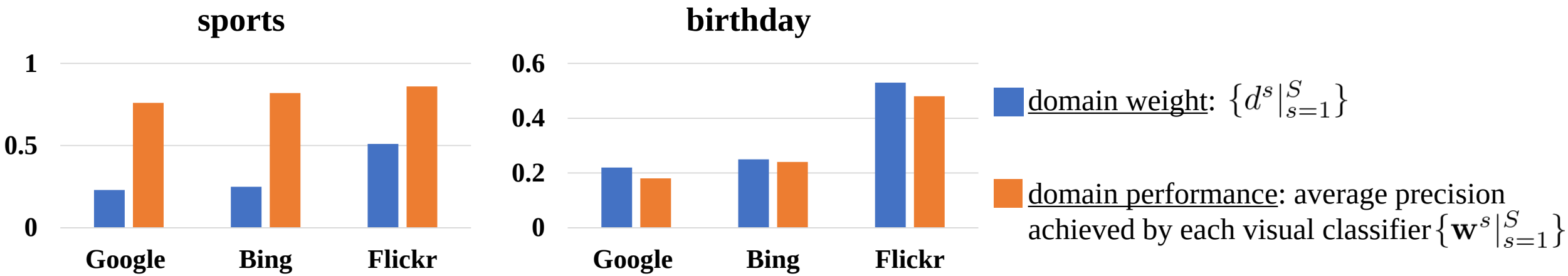
Experiments

Video Event Detection on CCV Dataset



Experiments

Video Event Detection on CCV Dataset



Mean average precision (%) on the CCV dataset

Domain adaptation baseline	Privileged information baseline	Composite baseline	Ours
62.19	61.55	63.45	**66.18**